

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 09:16:40 ; Search time 95.52 seconds
(without alignments)
6081.692 Million cell updates/sec

Title: US-09-923-831-42

Perfect score: 2365
Sequence: 1 ttgtacgcgagctcgatcc.....aaaaaaaaaaaaaaaa 2365

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCrus.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2365	100.0	2365	4 US-09-183-706-42	Sequence 42, Appl
2	2365	100.0	2365	4 US-09-567-993-42	Sequence 42, Appl
3	307.8	13.0	323	4 US-09-183-706-1	Sequence 1, Appl
4	307.8	13.0	323	4 US-09-567-993-1	Sequence 1, Appl
5	186.4	7.9	3825	4 US-09-208-742-3	Sequence 3, Appl
6	151	6.4	1254	5 PCT-US96-05320X-894	Sequence 894, Appl
7	143.4	6.1	2319	3 US-09-058-489-90	Sequence 90, Appl
8	143.4	6.1	4416	3 US-09-058-489-17	Sequence 17, Appl
9	135.4	5.7	3408	3 US-09-058-489-14	Sequence 14, Appl
10	135.4	5.7	5322	3 US-09-058-489-13	Sequence 13, Appl
11	120.6	5.1	1800	3 US-09-038-773A-1	Sequence 1, Appl
12	107.8	4.6	43676	3 US-09-356-952-12	Sequence 12, Appl
13	101.6	4.3	1682	4 US-09-318-443-7	Sequence 7, Appl
14	98.4	4.2	1536	4 US-09-318-443-5	Sequence 5, Appl
15	96.4	4.1	1245	4 US-09-318-443-1	Sequence 1, Appl
16	87	3.7	1347	2 US-08-959-749-1	Sequence 1, Appl
17	87	3.7	1347	2 US-09-351-497-1	Sequence 1, Appl
18	73.2	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
19	64.4	2.7	1374	1 US-08-928-738-2	Sequence 2, Appl
20	64.4	2.7	1614	3 US-09-008-271A-16	Sequence 16, Appl
21	58.2	2.5	2061	4 US-08-607-509-3	Sequence 3, Appl
22	55.8	2.4	1867	2 US-08-634-642-3	Sequence 3, Appl
23	55.8	2.4	1867	2 US-08-989-370-3	Sequence 3, Appl
24	55.8	2.4	1867	2 US-09-503-391-3	Sequence 3, Appl
25	53	2.2	1997	4 US-08-533-669A-9	Sequence 9, Appl
26	52.6	2.2	1618	2 US-08-607-509-1	Sequence 1, Appl
27	52.6	2.2	1618	2 US-08-607-509-1	Sequence 1, Appl

28	52.6	2.2	1618	2 US-08-454-036-1	Sequence 1, Appl
29	52.6	2.2	1618	2 US-08-634-642-1	Sequence 1, Appl
30	52.6	2.2	1618	2 US-08-989-370-1	Sequence 1, Appl
31	52.6	2.2	1618	5 PCT-US95-05064-1	Sequence 1, Appl
32	52.2	2.2	532	4 US-09-318-443-3	Sequence 3, Appl
33	51.2	2.2	2619	5 PCT-US96-10521-17	Sequence 17, Appl
34	51.2	2.2	2649	2 US-08-718-964-1	Sequence 1, Appl
35	51.2	2.2	2649	2 US-09-059-964A-1	Sequence 1, Appl
36	51.2	2.2	2649	2 US-08-842-341-1	Sequence 1, Appl
37	51.2	2.2	2887	5 PCT-US96-10521-14	Sequence 14, Appl
38	50.8	2.1	235	2 US-08-245-511-19	Sequence 19, Appl
39	50.8	2.1	235	2 US-08-600-993A-19	Sequence 19, Appl
40	50.6	2.1	1582	3 US-08-545-196B-10	Sequence 10, Appl
41	50.6	2.1	1582	3 US-08-545-196B-12	Sequence 12, Appl
42	50.6	2.1	2934	3 US-09-149-934-2	Sequence 2, Appl
43	50.2	2.1	342	2 US-08-245-511-17	Sequence 17, Appl
44	50.2	2.1	342	2 US-08-600-993A-17	Sequence 17, Appl
45	49.8	2.1	1656	4 US-09-503-391-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-09-183-706-42
; Sequence 42, Application US/09183706
; Patent No. 6245525
; GENERAL INFORMATION:
; APPLICANT: Martelange, Valrie
; APPLICANT: Boon-Pailleur, Thierry
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR
; FILE REFERENCE: 10461/7054
; CURRENT APPLICATION NUMBER: US/09/183,706
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: 09/122,989
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 42
; LENGTH: 2365
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (208)...(2151)
US-09-183-706-42

Query Match 100.0%; Score 2365; DB 4; Length 2365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtacgcgagctcgatccctagtaacgcgcgcagtgctgtaaaagttggcgcgga 60
I ttgtacgcgagctcgatccctagtaacgcgcgcagtgctgtaaaagttggcgcgga 60
Db 1 ttgtacgcgagctcgatccctagtaacgcgcgcagtgctgtaaaagttggcgcgga 60
QY 61 taagagcgtgagcgaggggagctagcctctgctggggctccttaagtaacgcgcgcgc 120
I taagagcgtgagcgaggggagctagcctctgctggggctccttaagtaacgcgcgcgc 120
Db 61 taagagcgtgagcgaggggagctagcctctgctggggctccttaagtaacgcgcgcgc 120
QY 121 ttccctgagcgtgcttcttcagcgtcaggtggtgacagcttgagcgcgcgcgc 180
I ttccctgagcgtgcttcttcagcgtcaggtggtgacagcttgagcgcgcgcgcgcgc 180
Db 121 ttccctgagcgtgcttcttcagcgtcaggtggtgacagcttgagcgcgcgcgcgcgc 180
QY 181 acgtcgagcgcgcgcctcttcgaaacaatgtccacacagagagagcctcccaagcctct 240
I acgtcgagcgcgcgcctcttcgaaacaatgtccacacagagagagcctcccaagcctct 240
Db 181 acgtcgagcgcgcgcctcttcgaaacaatgtccacacagagagagcctcccaagcctct 240
QY 241 acgtggtgctgttcctcagtcgcgcgaagctgcgaagtgctccgagcgcgcgaagagagcg 300
I acgtggtgctgttcctcagtcgcgcgaagctgcgaagtgctccgagcgcgcgaagagagcg 300
Db 241 acgtggtgctgttcctcagtcgcgcgaagctgcgaagtgctccgagcgcgcgaagagagcg 300
QY 301 gcgagagagtgtaacgcacaggtccctgagagatatagtgctgcgaagagtggtcgtcg 360
```

Dh	1381	tacttctgttttagatctggaagcagaacagatgcttgcatacttggaacctgaacctgaatg	1440
Qy	1441	aagatatttgtagatctgctgccccagcttagcgcagccagttcttgcacagtgctctacagccct	1500
Dh	1441	aagatatttgtagatctgctgccccagcttagcgcagccagttcttgcacagtgctctacagccct	1500
Qy	1501	cattcagcttcatactgcctctgcacaaacttcttatgaaagaaacaaatgatctgtctatgtgt	1560
Dh	1501	cattcagcttcatactgcctctgcacaaacttcttatgaaagaaacaaatgatctgtctatgtgt	1560
Qy	1561	accttggatctagtgtctgtcaagtctcaatgtagaagcacaatatatcttgaacccgagga	1620
Dh	1561	accttggatctagtgtctgtcaagtctcaatgtagaagcacaatatatcttgaacccgagga	1620
Qy	1621	gagaaaaatgtagtccatctgcaaaacttcttcacagagtaatgcatccatccacagaacagctc	1680
Dh	1621	gagaaaaatgtagtccatctgcaaaacttcttcacagagtaatgcatccatccacagaacagctc	1680
Qy	1681	gtcttcgtttcttcgaaaagcctgtcttggagctcctctcaaatgtaaccttaattcttggaa	1740
Dh	1681	gtcttcgtttcttcgaaaagcctgtcttggagctcctctcaaatgtaaccttaattcttggaa	1740
Qy	1741	atactcagtagagctctctctgcatctgtagagaatgaagaacagagagatctgggagaaagcatlag	1800
Dh	1741	atactcagtagagctctctctgcatctgtagagaatgaagaacagagagatctgggagaaagcatlag	1800
Qy	1801	aactcttaaaaaacggagaaagctgtagaatactcaacttgaacttgaactgaactgaactcttagagact	1860
Dh	1801	aactcttaaaaaacggagaaagctgtagaatactcaacttgaacttgaactgaactgaactcttagagact	1860
Qy	1861	gattctcactgacgcttaccacatgctctataaattcttgactcttcacaggaatatctgaagataac	1920
Dh	1861	gattctcactgacgcttaccacatgctctataaattcttgactcttcacaggaatatctgaagataac	1920
Qy	1921	gtacacccggaataggcgacagcgagaaagcagagagacttggctgttccattacaaacttgg	1980
Dh	1921	gtacacccggaataggcgacagcgagaaagcagagagacttggctgttccattacaaacttgg	1980
Qy	1981	actagaatactgattggagaggtctgctctgtaattgatatcttcggaagagcgaaataag	2040
Dh	1981	actagaatactgattggagaggtctgctctgtaattgatatcttcggaagagcgaaataag	2040
Qy	2041	agtaatccagagagagactgtaatcgaatctgacatgagagagctcttgagcgacataccaaacgaaag	2100
Dh	2041	agtaatccagagagagactgtaatcgaatctgacatgagagagagctcttgagcgacataccaaacgaaag	2100
Qy	2101	gaaatctggaaagaaaaatggaagaagcctccaagaaagcgccaaagaagttcttaataatgtct	2160
Dh	2101	gaaatctggaaagaaaaatggaagaagcctccaagaaagcgccaaagaagttcttaataatgtct	2160
Qy	2161	ctgtactacagtgaggctgtagagaaatctcaagaattcttgaagaataatgtaagaacagagatctg	2220
Dh	2161	ctgtactacagtgaggctgtagagaaatctcaagaattcttgaagaataatgtaagaacagagatctg	2220
Qy	2221	gactctgttgcagatctgaagaagcaggagacttggactgtcttcttaaaataatgtagtct	2280
Dh	2221	gactctgttgcagatctgaagaagcaggagacttggactgtcttcttaaaataatgtagtct	2280
Qy	2281	tgaaaaatatagaatccagtgctttatatacttcttcttaataaaatagaagatatttaacct	2340
Dh	2281	tgaaaaatatagaatccagtgctttatatacttcttcttaataaaatagaagatatttaacct	2340
Qy	2341	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa	2365
Dh	2341	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa	2365
RESULT 2			
US-09-567-995-42			
: Sequence 42, Application US/09567995			
: Patent No. 6303756			
: GENERAL INFORMATION:			
: APPLICANT: Martelange, Val,rie			

```

RESULT 2
US-09-567-995-42
; Sequence 42, Application US/09567995
; Patent No. 6303756
; GENERAL INFORMATION:
; APPLICANT: Martelange, Val, rle

```

APPLICANT: De Smel, Charles
APPLICANT: Boon-Failleur, Thierry
TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR
FILE REFERENCE: L0461/7054
CURRENT APPLICATION NUMBER: us/09/567,995
PRIORITY FILING DATE: 2000-05-10
PRIORITY APPLICATION NUMBER: 09/183,706
PRIORITY FILING DATE: 1998-10-30
NUMBER OF SEQ ID NOS: 43
SEQ ID NO 42
LENGTH: 2365
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (208)...(2151)
US-09-567-995-42

Query Match 100.0%; Score 2365; DB 4; Length 2365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttgtaccagagctggatccctagtaagagccgacgtgtgtgtgtaaaagtggcgga 60
D 1 ttgtaccagagctggatccctagtaagagccgacgtgtgtgtgtaaaagtggcgga 60
QY 61 taagagcgctggcgggggtctagctctgtcgggctccttaagtaagcgtgtgctgc 120
D 61 taagagcgctggcgggggtctagctctgtcgggctccttaagtaagcgtgtgctgc 120
QY 121 ttccctgacagctagctcttaagacgttcacagtggtgtgtcagaagcttgagcgcaag 180
D 121 ttccctgacagctagctcttaagacgttcacagtggtgtgtcagaagcttgagcgcaag 180
QY 181 acgtcggagcgccctctctgtgaaacaatgtccaccagagagagctcccaagccctct 240
D 181 acgtcggagcgccctctctgtgaaacaatgtccaccagagagagctcccaagccctct 240
QY 241 acgtgtgtctgtctagctcgtgcgaagctcgacaagtgctcccgagcgcgaaagagcgcg 300
D 241 acgtgtgtctgtctagctcgtgcgaagctcgacaagtgctcccgagcgcgaaagagcgcg 300
QY 301 gcggaagagctgaatcgaacaggtccctgaaggaataagtgctcgagaggtgtcgctgg 360
D 301 gcggaagagctgaatcgaacaggtccctgaaggaataagtgctcgagaggtgtcgctgg 360
QY 361 agagggcaccttaaggcccggaagcggtgctgtgttcacgaagaaactgtccgttgt 420
D 361 agagggcaccttaaggcccggaagcggtgctgtgttcacgaagaaactgtccgttgt 420
QY 421 ttgtcttgaagagccaacttgt 480
D 421 ttgtcttgaagagccaacttgt 480
QY 481 aatatacaagtaacaacaacaacaaatccaataatacaagaacaacgaataatca 540
D 481 aatatacaagtaacaacaacaacaaatccaataatacaagaacaacgaataatca 540
QY 541 gtcaaaattttgtgcaagaagcatgtcaaacgaagaacgaacgagtataagacaattt 600
D 541 gtcaaaattttgtgcaagaagcatgtcaaacgaagaacgaacgagtataagacaattt 600
QY 601 gttaaaagctagaagaataatataatcagaatcgcgaatctgatactgtcaacct 660
D 601 gttaaaagctagaagaataatataatcagaatcgcgaatctgatactgtcaacct 660
QY 661 tctgttgaaaaagatggaagcacagataaacaatgtgtgtgcagagatcgcgcatgtata 720
D 661 tctgttgaaaaagatggaagcacagataaacaatgtgtgtgcagagatcgcgcatgtata 720
QY 721 gattggatccaatagaaggggaagtgttgaatgtgcaaaaacaaagtggcagattca 780
D 721 gattggatccaatagaaggggaagtgttgaatgtgcaaaaacaaagtggcagattca 780

D 721 gattggatccaatagaaggggaagtgttgaatgtgcaaaaacaaagtggcagattca 780
QY 781 ccaccaatagaaaaaactttataaaggtccatgcgcacgaagtgcatagtcaaatga 840
D 781 ccaccaatagaaaaaactttataaaggtccatgcgcacgaagtgcatagtcaaatga 840
QY 841 gaagcagaatgtggaagaaagaaatttaataaagtggtgagatgacttgaagagtg 900
D 841 gaagcagaatgtggaagaaagaaatttaataaagtggtgagatgacttgaagagtg 900
QY 901 gagaagacactatcccccaatctctacgtgcaatttgaatgagccttcaagtatcc 960
D 901 gagaagacactatcccccaatctctacgtgcaatttgaatgagccttcaagtatcc 960
QY 961 gaggctatgaaaaacatlaaaaaggcaggtttcaaaaagccaacacatctacgacag 1020
D 961 gaggctatgaaaaacatlaaaaaggcaggtttcaaaaagccaacacatctacgacag 1020
QY 1021 gcatgcccattgtgtgtgcaaggaatagaatcttaagagtgagccagactggaacaga 1080
D 1021 gcatgcccattgtgtgtgcaaggaatagaatcttaagagtgagccagactggaacaga 1080
QY 1081 aagacattgttataatgaatcgtgatttcaatcgtgtccctcaacccagccttaaa 1140
D 1081 aagacattgttataatgaatcgtgatttcaatcgtgtccctcaacccagccttaaa 1140
QY 1141 ggtcaaaagaaatgacccggcagtgtaagttaactccactgagggaattagcatttaa 1200
D 1141 ggtcaaaagaaatgacccggcagtgtaagttaactccactgagggaattagcatttaa 1200
QY 1201 gtaagaaggaatgttgcataatctcaataaagggtctcgaggtgtgtgtgtatagt 1260
D 1201 gtaagaaggaatgttgcataatctcaataaagggtctcgaggtgtgtgtgtatagt 1260
QY 1261 ggtggaataagaatgaaacaataatagaagagcttaaaaagggtgagatatacaaatga 1320
D 1261 ggtggaataagaatgaaacaataatagaagagcttaaaaagggtgagatatacaaatga 1320
QY 1321 actccggaaagttaagatctgcgaatgtgaactgtgtcaactgtgaagaataaac 1380
D 1321 actccggaaagttaagatctgcgaatgtgaactgtgtcaactgtgaagaataaac 1380
QY 1381 tacttggttttaagaatgaagcagaagaatgtgtgaacatgttggaatcgaacgaata 1440
D 1381 tacttggttttaagaatgaagcagaagaatgtgtgaacatgttggaatcgaacgaata 1440
QY 1441 aagatttgttaagt 1500
D 1441 aagatttgttaagt 1500
QY 1501 catcagatctacgtccctgcgaacatcttaattgaaagaacaaatgtgtgtgtgtgt 1560
D 1501 catcagatctacgtccctgcgaacatcttaattgaaagaacaaatgtgtgtgtgtgt 1560
QY 1561 acattggaatgaatgt 1620
D 1561 acattggaatgaatgt 1620
QY 1621 gagaatgtgaatcagatcaaacattttcaacaggtgtgtgtgtgtgtgtgtgtgtgt 1680
D 1621 gagaatgtgaatcagatcaaacattttcaacaggtgtgtgtgtgtgtgtgtgtgtgt 1680
QY 1681 gttctgtttctcgaaaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740
D 1681 gttctgtttctcgaaaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740
QY 1741 atatacagtaagctctgt 1800
D 1741 atatacagtaagctctgt 1800
QY 1801 aactttaaacagcgaagtgtgaataactaatgtcaactgtgacttgaagccttagagact 1860
D 1801 aactttaaacagcgaagtgtgaataactaatgtcaactgtgacttgaagccttagagact 1860

```
OY 1861 gatgtccatgacgtttacaatgtctataatttgaactttccacggagaatttgaagaatc 1920
|||||
DB 1861 gatgtccatgacgtttacaatgtctataatttgaactttccacggagaatttgaagaatc 1920
OY 1921 gtacccgaataaggcgacacggagaagcagagagctgtgtttccattcaactctg 1980
|||||
DB 1921 gtacccgaataaggcgacacggagaagcagagagctgtgtttccattcaactctg 1980
OY 1981 actggaattgattggaggggttgcctctgaattgaatttaatttcttgagaagaacaatcag 2040
|||||
DB 1981 actggaattgattggaggggttgcctctgaattgaatttaatttcttgagaagaacaatcag 2040
OY 2041 agtattccagaaggagcttctatcaatggctcgaagaggttctgaggcacatcaacygaanaag 2100
|||||
DB 2041 agtattccagaaggagcttctatcaatggctcgaagaggttctgaggcacatcaacygaanaag 2100
OY 2101 gaaatggagaagaaatctgaaagacctcaaggagggccgaagaagtttcaatgaagctt 2160
|||||
DB 2101 gaaatggagaagaaatctgaaagacctcaaggagggccgaagaagtttcaatgaagctt 2160
OY 2161 cctgtactagttgggttagagaatccaagaatttctttagaataatagtagaagacagaaglatcg 2220
|||||
DB 2161 cctgtactagttgggttagagaatccaagaatttctttagaataatagtagaagacagaaglatcg 2220
OY 2221 gacatgttgcagatgaaagagacggagctgattgtgctatctcttaataataatagtggt 2280
|||||
DB 2221 gacatgttgcagatgaaagagacggagctgattgtgctatctcttaataataatagtggt 2280
OY 2281 tgaataataagatccacagtgcttcttaacttctttaaataaataagatatttaactt 2340
|||||
DB 2281 tgaataataagatccacagtgcttcttaacttctttaaataaataagatatttaactt 2340
OY 2341 aaaaaaaaaaaaaaaaaaaaaa 2365
|||||
DB 2341 aaaaaaaaaaaaaaaaaaaaaa 2365
```

```
RESULT 3
US-09-183-706-1
; Sequence 1, Application US/09183706
; Patent No. 6245525
; GENERAL INFORMATION:
; APPLICANT: Martelange, Val'rie
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR
; FILE REFERENCE: 10461/7054
; CURRENT APPLICATION NUMBER: US/09/183.706
; CURRENT FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: 09/122,989
; EARLIER FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 1
; LENGTH: 323
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-183-706-1
```

```
Query Match 13.0%; Score 307.8; DB 4; Length 323;
Best Local Similarity 98.8%; Pred. No. 2.7e-69;
Matches 321; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

OY 727 gatcaaatctagagagaggttgaatctgcaaaaacaaagtgaggcagattaccacca 786
|||||
DB 1 gatcaaatctagagagaggttgaatctgcaaaaacaaagtgaggcagattaccacca 60

OY 787 attagaagaaactttataaagagctccactgcccaagtgccatgtgcaaaagtagaagca 846
|||||
DB 61 attagaagaaactttataaagagctccactgcccaagtgccatgtgcaaaagtagaagca 120

OY 847 gatagttggaggaagaaattttaataacgttggatgacttgaaagatgggagaa 906
```

```
|||||
DB 121 gatagttggaggaagaaattttaataacgttggatgacttgaaagatgggagaa 180
|||||
OY 907 cgaacctaccacactctactctgacatttgatgacgccttcaatgttatccctgaagtt 966
|||||
DB 181 cgaacctacc--aatctactctgacatttgatgacgccttcaatgttatccctgaagtt 238
OY 967 atgaaacatttaaaaagcaggttctcaaaagccacacactatcactgaagcagctg 1026
|||||
DB 239 atgaaacatttaaaaagcaggttctcaaaagccacacactatcactgaagcagctg 298
OY 1027 cccattgtgtcgaaggaatagatc 1051
|||||
DB 299 cccattgtgtcgaaggaatagatc 323
```

```
RESULT 4
US-09-567-995-1
; Sequence 1, Application US/09567995
; Patent No. 6303756
; GENERAL INFORMATION:
; APPLICANT: Martelange, Val'rie
; APPLICANT: De Smet, Charles
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: TUMOR ASSOCIATED NUCLEIC ACIDS AND USES THEREFOR
; FILE REFERENCE: 10461/7054
; CURRENT APPLICATION NUMBER: US/09/567.995
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/183.706
; PRIOR FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 1
; LENGTH: 323
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-567-995-1
```

```
Query Match 13.0%; Score 307.8; DB 4; Length 323;
Best Local Similarity 98.8%; Pred. No. 2.7e-69;
Matches 321; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

OY 727 gatcaaatctagagagaggttgaatctgcaaaaacaaagtgaggcagattaccacca 786
|||||
DB 1 gatcaaatctagagagaggttgaatctgcaaaaacaaagtgaggcagattaccacca 60

OY 787 attagaagaaactttataaagagctccactgcccaagtgccatgtgcaaaagtagaagca 846
|||||
DB 61 attagaagaaactttataaagagctccactgcccaagtgccatgtgcaaaagtagaagca 120

OY 847 gatagttggaggaagaaattttaataacgttggatgacttgaaagatgggagaa 906
|||||
DB 121 gatagttggaggaagaaattttaataacgttggatgacttgaaagatgggagaa 180

OY 907 cgaacctaccacactctactctgacatttgatgacgccttcaatgttatccctgaagtt 966
|||||
DB 181 cgaacctacc--aatctactctgacatttgatgacgccttcaatgttatccctgaagtt 238
OY 967 atgaaacatttaaaaagcaggttctcaaaagccacacactatcactgaagcagctg 1026
|||||
DB 239 atgaaacatttaaaaagcaggttctcaaaagccacacactatcactgaagcagctg 298

OY 1027 cccattgtgtcgaaggaatagatc 1051
|||||
DB 299 cccattgtgtcgaaggaatagatc 323

RESULT 5
US-09-208-742-3
; Sequence 3, Application US/09208742
; Patent No. 6174679
; GENERAL INFORMATION:
; APPLICANT: Kaufmann, Joerg
```

Thu Jul 25 15:15:58 2002

us-09-923-831-42.rni

Page 5

? TITLE OF INVENTION: CIP150/hnRf1150 Is Necessary for Cell
 ? TITLE OF INVENTION: Cycle Progression
 ? FILE REFERENCE: 1453.002
 ? CURRENT APPLICATION NUMBER: US/09/208,742
 ? CURRENT FILING DATE: 1998-12-10
 ? NUMBER OF SEQ ID NOS: 6
 ? SOFTWARE: FastSeq for Windows Version 3.0
 ? SEQ ID NO. 3
 ? LENGTH: 3825
 ? TYPE: DNA
 ? ORGANISM: human
 ? US-09-208-742-3

Query Match	7.9%;	Score 186.4;	DB 4;	Length 3825;
Best Local Similarity	49.38;	Pred. No. 6.3e-38;		
Matches 546;	Conservative	0;	Mismatches 556;	Indels 6;
				Gaps 2;

QY	965	tttttggaaaactttaaaaggccggttttccaaaagccaaaccctatttcagtcacggatc	10234
Db	926	ttatgcaaccaggtcttcggaatctggaatacaacaagccaccctccaaatacagtcagagtg	985
QY	1025	ggcccatctgtctgcagaagatagatcttatagtgatggatcccaactcggaaacgggaaga	10844
Db	986	tgcctgtgacctaaagtggtgagagacatgattgttatctgcacaaaacgggtagtgggaa	10455
QY	1085	catgtgtatttatatgctcggaatttatccatctgttccttcaaccacgcttaaggtc	11444
Db	1046	ctgcagcctccatcttgcgccagtgctgattccataataatgagcamaagagattgaa---c	11020
QY	1145	aaagagatagaaccgggacatctagttcttaaccctccacccggaaattagcacttcaatag	12004
Db	1103	caagtatgagcaacaaatctgcagtgattgtgtctctcaaccaggagcttgcacgagatcc	11622
QY	1205	aagggaaagtgtgcgaataattctac---aaggcctcggagtggttggatatggtg	12611
Db	1163	atcccaaatgtaagcgggtcttggaagaagcataatcttcgactcagtcggttatgtagag	12222
QY	1262	gtggaatatgagatgaaacaataagaaagcttaaaaagaagtgtagatatcatatgcaa	13211
Db	1223	gaaggaggtatgtggygacagggccaaagcccttcagggaggggcagagattgtgtgtga	12822
QY	1322	ctcccggaagattgaatgactgcgaatgagtaacttcctgtaaacttgaagataataact	13811
Db	1283	cccccaagctcgaactgatatgacacgtgaaaaagaagctaaccaattccaaagagctctc	13422
QY	1382	acttggttttaagatgaagcagacaagaatgtctggaacatltggaattgaaacccaataatga	14411
Db	1343	accttggttttgaatgaagcagacatgaaatgtctgtgacatgtgattgtgattaccaagttcat	14020
QY	1442	agatttgttgatgtagtgcgcccgaatgaagcagaagctatgaccagtgctaaatgccc	15011
Db	1403	ccatagcaagatcatgctgcgtcccgaaagcggaacccctcttataagtgacaacttctcgga	14622
QY	1502	atcagttcatctgcctcgcgacaactcttaatttgaagaacaaatgatgtgtctatgtgtga	15611
Db	1463	agaaagattgaaaagttgccaagagacatccctgatccagccattctcgagtgtgcagggg	15222
QY	1562	catgtgacttaatttcgtctgaatgtcagtgaaagcaaaatataattgtataccaacggagga	16211
Db	1523	atattgagagagcaaatggaagatgtgacagaaattgtgagatctccattcttggaacta	15822
QY	1622	agaaatvgagatcacagtgcaaaactttctacagagtatgtcatccacagacaagaattcatg	16811
Db	1583	gtaaatvgagaaactgtcttaaccggcgtctgttagaatttaactcttcagggagtgtccctc	16422
QY	1682	ttctgtttcttcgaaaagcctgtgtcggaatcacatttaagtgaaacttaacttggaaata	17411
Db	1643	ttctgttctctaaaaagccaatgctcggaagagctcgtcgaaataaacttaacaggaaggtc	17022
QY	1742	tatcagtagagctcttcgatgtgagatagaaacagagatctcggagaaagcatagaga	18011
Db	1703	ataatcttgcgtctgcatactgagatgatgtacaaagtgtgaagaacaagaatcatctacg	17662

QY	1802	actctaaacccgagcaagtagatgatactcattgacagctgacgtactagctctcaagagactg	1861
Db	1763	actcttaagaaaagagacatcccgctccgtgtagccacagatgtgcagcccgtygtctg	18222
QY	1862	atgtccatgaagcttacacagctctataatttgactctccacggatattgaagaatag	1921
Db	1823	acattctccccaattagaagactgctcattcaactaagaatgtagccggagaaatttgatacgcaca	1882
QY	1922	tacaccgaatagagcgccacagggagagcgagagactgtagtcttcattcaacttga	1981
Db	1883	ctcacagagcttgcgcgcacaggaagagcggtgtagaaaaggtgtggtccatactactac	1942
QY	1982	ctagaagaatgatgagaggttgcctctcgaattatcatatctctggaagaagcgcaatcaga	2041
Db	1943	ctccccaagacagacatttctctgtgacacctgtgtccggaacttggaaagcgccaatcaac	2002
QY	2042	gtaattccagagagctgtgatactatgac	2069
Db	2003	acgtttctaaagaactcctcagatctgc	2030

```

RESULT 6
PCT-US96-05320A-894
: Sequence 894, Application PC/TUS9605320A
:
: GENERAL INFORMATION:
: APPLICANT: Human Genome Sciences
: APPLICANT: 9410 Key West Avenue
: APPLICANT: Rockville, MD 20850
: APPLICANT: United States of America
: APPLICANT: Johns Hopkins University
: APPLICANT: 720 Rutland Avenue
: APPLICANT: Baltimore, MD 21205
: APPLICANT: United States of America
: APPLICANT: Mark D. Adams
: APPLICANT: Owen White
: APPLICANT: Hamilton O. Smith
: APPLICANT: J. Craig Ventner
: TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genom
:
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20003-3934
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/05320A
: FILING DATE: April22, 1996
:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/476,102
: FILING DATE: June 7, 1995
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/487,429
: FILING DATE: June 7, 1995
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Eric K. Steffe
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1488.014PC01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
:
: INFORMATION FOR SEQ ID NO: 894:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1254 base pairs
: TYPE: nucleic acid
:

```

STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-894

Query Match 6.4%; Score 151; DB 5; Length 1254;

Best Local Similarity 49.6%; Pred. No. 3,9e-29;

Matches 495; Conservative 0; Mismatches 495; Indels 7; Gaps 4;

```
QY 981 aaagcaggtttccaagaacacacctaattcagtcacaggtcaggtccattgtgtgca 1040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 AAAAAAGCGTTTATTTGTTGTTACCCCAATTCAGCCTTATCCCTGCTTATCCGTTAA 143

QY 1041 aggaatagatcttatagagtagccagactggaacaggaagacattgttattaat 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 TGGACGAGATGTCCAGACMACCTCAAACTGTACAGCGCAAGCAATGGCTTTTAAAC 203

QY 1101 gccctgatttcatcctgtcctcctcaaccagacctaaagtcagaagatagaccgg 1160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 GGCCTACTTTTACCATCTTTTACTCACCAGATCCTTAATCTTAATATCCACCCAAAG 263

QY 1161 catgttagttcctaacctccactcgggaattagacttaagt--agaaggaaatgtg 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 ACCTTTGATTTTACCACTTACGCAATTCAGCAATTAACGGTACAGATTAGTATCCGCAAT 323

QY 1218 caaatattcatataaagggtctcgaggtttgtgtatagtgtgtgaatatagatga 1277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 TCTTGCAAAAGCGAGTGGATTAAAGACCGCACTTGCCATGCTGCGCATGCTTATGATTA 383

QY 1278 acaaatagaagagcttaaaaaagtgtagatatacatacttgcaactccggaaagtga 1337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 ACAACTACAAAGCATTTGAGCGTGCCTCGATTTTGTGATGTTGTTGTTACGCGGGGAGTCA 443

QY 1338 tgatctgcaaatgagtaactctgcaatctgaaagataaactacttggttttagatga 1397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 TCATTATGTGAACAACGCGTAAATGCTTTAGATGAATCCAACTTGCTGCTGATGATGA 503

QY 1398 aagagacaagagtgtagacatgtagattggaacccagaaatgaagaatttgttagagt 1457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 ACCAGATGGAATGTTTATCTTGCGTT--ATCCGCTGATTTATCTTTATTTATGCGTAAA 561

QY 1458 gcgcccagatagagaacgtatgaccagtgctacatgagcctcatcagttcatcgct 1517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 TGCCTCCGCTCCGCAAGCTCGTTAACGATGTTATTTTACGAGACGCTTCTTAAAGTG 621

QY 1518 gcacaaatctatttgaagaacaaatgtagtctatgtgtgtagcatitgagatcgtgc 1577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 622 CGTGAATTAGCATTTGAAAGTATGAAATGAACTGAAATATTTGAAATGAAACAGAAACA 681

QY 1578 tga-agltcagtgaaagcaaatataatgtlaaccacaggaagaaagaaatggaatgaca 1636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 682 AAAACAGGGCACCGAATTTAAAGAAACTTTTATCCATCTATTCAGATAAATGGCA 741

QY 1637 tgaacactttctcagagatgt-catccacagacaagtcattgtctgtcttcctga 1695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 CTTCCTTAACCTTAATGGAAGTGAATGCTGACGCTGTATTGATTTGGAAATACG 801

QY 1696 aaagctgttcggaatcactatcaagtgacctaacttggaatatatcatcagtagagtc 1755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 AAACATCGTTGTGAAGAAATTTGGGCTATTGCGGCTGATGGGCACTCGTGGGTTTA 861

QY 1756 ctgcactggaatagagaacagagaatgcgggaaagaagatagaagaaactttaaaccagc 1815
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 CTGACTGGCGATGAGCAGACAGAAAAACCTTATTCGTTATTTAAACAAATTTACTGATGG 921

QY 1815 aaagtgaataactaatltgcaactgtatagcctctagagacttgatgtcatagcgt 1875
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 GATTTGATATTATTAGTGGCAACAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 981

QY 1876 acaatgctataatlttagacttccaacgaaatltgaagaatacgtaaaccgagatag 1935
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 982 ACGCATGTTTCAATTTATGATTTACCCGATGATCGAAGAAATTTATGTTACCGAATTTGG 1041
```

```
QY 1936 cgcacggaagagacagagagactggtcttccatla 1972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 CGTACTGACAGACAGACGGAAGTGTGTTCCATTTA 1078
```

RESULT 7

US-09-058-489-90

; Sequence 90; Application US/09058489

; Patent No. 6103886

; GENERAL INFORMATION:

; APPLICANT: Whitehead Institute for Biomedical Research

; APPLICANT: Lahn, Bruce

; APPLICANT: Page, David

; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

; FILE REFERENCE: WHI97-08PA

; CURRENT APPLICATION NUMBER: US/09/058,489

; EARLIER FILING DATE: 1996-04-10

; EARLIER FILING DATE: 1997-04-11

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: FastSeq For Windows Version 3.0

; SEQ ID NO 90

; TYPE: DNA

; ORGANISM: Human

US-09-058-489-90

Query Match 6.1%; Score 143.4; DB 3; Length 2319;
Best Local Similarity 49.7%; Pred. No. 4,4e-27;
Matches 459; Conservative 0; Mismatches 446; Indels 18; Gaps 3;

```
QY 1165 ttggtcttaactccactcgggaattagcaactcaagttagaaggaagatgtgcaaat 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 ttggttttagccccaagaagatgtgcgtacagatctatgaggaagccaagaattt 922

QY 1225 tcaataaa---ggcttcggaagtgtgtgtatagtggtggtggaataagagaaacaa 1281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 tcttaacgactagatgttcgtctcctgtgtatgtatgtgtgtgtgtgtgtgtgtgtgt 992

QY 1282 atagaagagctttaaanaaaggctgtagatatacataatgtcaaccctccggaagat 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 atcgggacttagaacgtgtagatgcacattgttagtagcaccctcagagcgtctagat 1052

QY 1342 ctgcaaatgagtaactctgtcaactcgaagaataataactactgtgttttagaagaa 1401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1053 atgatggaagaggaagaatcttgatctgaactctcgaagtaacttagtgttgatgaa 1112

QY 1402 gacaaagtcttggaatggaattggaacccagaataatgaagatttgttagatc----- 1456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 gatagagtgatgatatgtgatttgaaacctcagatacgtcgtatagttggaacaagata 1172

QY 1457 -----tgcgccagatagcagagatataatgaacagtgtaacttgccctatcagtt 1509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1173 atgcacacaaaggcgttcgtcaacacaaatgatttagtagtaactttccaaagaata 1232

QY 1510 catgcctcgcgaacactctatttgaagaacaaatgattgtcatgtgtgtacattgagat 1569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1233 cagatgtcgtcgtgacttttgatgcatatatacttttgcgtaggaagagtaggc 1292

QY 1570 ctggtgtgtgtagttcagtggaagcaaatataatgtgtatcaacacggaaggaagaa 1629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1293 tctactctctgaga---acatcaacacagaagaagtttgggtggaagacttagaaacgg 1349

QY 1630 agtacaatgcaactttctacaagatgtgtcatccacagaaagaagatctgtcttgt 1689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1350 tcatcttactgtgacatttgatgcaacaggaagtgatctcaacttacttagttgtgt 1409

QY 1680 tctcgaagaatcgtgtggtggaacttatacaagtgacttaatacttggaataatactg 1749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1410 gagacaaaagaagagacagattcccttgagagattcttataccatgaaagatatgtctgt 1469
```

Qy	1750	gagctctgcgacgagatagaagaacacggaagatctcgggagaagcatatagaacatttaa	1809
Db	1470	actagatcatcagagagacggtccacgagagatcgaagagagggcccttaaccggatttcgc	1529
Qy	1810	acacgacgaagctgagatatacctaattgcaaacgtgatacttagagagacttgatccat	1869
Db	1530	tccgagaaaagagcccaattctagctggtctacagctctgycagacagagatcagacattta	1589
Qy	1870	gagcttaccacatgcttaataatttgaccttcacaggaatatgtagaagatacgtaaccca	1929
Db	1590	aatgctgagacatgataccaatttgcattttgcgaatgtaattgaagatattgycatcgt	1649
Qy	1930	ataggcgacagcggagaagacagcggagagactggtgttccattacaacttgcatagaat	1989
Db	1650	atggcgcgacagcgacgctgtatggaaaacctgggaccttgcacacctatctttaaataaa	1709
Qy	1990	gatttgaaggtgtgacctctgcaattatattcctcggaaagacaaatcagatattcca	2049
Db	1710	aataatgaataatacacaagatttgttgatattcttcttgtagaagctaaacaagagtgct	1769
Qy	2050	gagagactgtatatacattgctgta	2072
Db	1770	tcttggcttgaaaataatgcttca	1792

```

RESULT      8
US-09-058-489-17
; Sequence 17 Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahm, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ. ID NOS.: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 4416
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-17

```

Query Match	6.18;	Score 143.4;	DB 3;	Length 4416;
Best Local Similarity	49.78;	Pred. No. 5.9e-27;		
Matches 459;	Conservative	0;	Mismatches 446;	Indels 18;
				Gaps 3;

QY	1165	tgtgtcttaccacccacccgaggaattagacactcaagtagaagaagatgttcgaacat	1224
Db	873	tgtgtttttgtccccaacaaagaatttgcgtctacagatctatgagsgaagccgaataatt	932
QY	1225	tcaatacaaa---ggccttcggagatgtttgtgtcatctgtgttgtaaaatagaatgtaca	1280
Db	933	tccacacgatacttagatgtctccctctgtcatgtttatagatgvgctgtcatctgtccagcg	992
QY	1282	atcagaagaagcttaaaaaagcgtgtcatatcatalcataaattgcacactcccggaagatgtgaat	1341
Db	993	atttcggagacttgaacgctgtgatacttctgttagtagcactccaagaacgtccatgtgact	1052
QY	1342	ctcgcaaatgagtaaacatttgtcaatctgaagaaataaacctactctgttttttagtagaaga	1401
Db	1053	atgatctggaagaagsgaagaatattggaatttggaacttttcgaagtacttagtctgttgtagact	1112
QY	1402	gacacagatcttggaacatggaattgtgaaccccaataaagaagaatttgttagatg----	1456
Db	1113	gatcggatctctgataatgtggaatttggaacctcagatcgtctatagtttgacaagaatact	1172

QY	1457	-----tgcgccagtagtcgagacaagctatagccagatgctacacagctcattcagtt	1509
Db	1173	atgcaccaaaagggctctgcacaccacatgcttttagtgtaattcttcccaagaata	1232
QY	1510	cacgcctctgcacaactctctctgaaagaaccaatgatgtctatgttgcatacttgat	1569
Db	1233	caagatctgtgcctcgactcttcttgcgtatatactcttctgtccttgaagcagaatgac	1292
QY	1570	ctagttgcgttaagttcagtcgaaagcaaatataatctgaaaccacgaggaagaatgag	1629
Db	1293	tcctactctggaga---acatcacacagaagtagtttgggtggaagaacttagataacgg	1349
QY	1630	agtcacatgcnaactttctacagagtatgtcatccacagacaaagtcatgtcttcgt	1689
Db	1350	tcaattctactctgacacattcttagatctgcnaacagggatgtaattcaattatgttttgg	1409
QY	1690	tctcgaaaagcgtgttcgcgaatcactatcaatgagccataacttgcgaatatatcagta	1749
Db	1410	gagacccaanaagggagcgaattccctcgggagttctctataccatgaaagtatgtctgt	1469
QY	1750	gagctctctcatgagatagagaacagagatctggaggaagcattagaagaaactttaa	1809
Db	1470	actagattctcatgtagagaccggtcaacagaaagatctgaaggaggcccttccacagttt	1529
QY	1810	acagagcaaaagtgcagatatcctaattgcacactgatactcagcctctaaaggaacttgat	1869
Db	1530	tcaagaaaagcccaattcttagtctgctatcagctctgtagcagacaggaactagacattca	1589
QY	1870	gaagctacacatgctataattttgacttccacaggaatatgaaagatactgcgaacgga	1929
Db	1590	aaatgtagagacatgctataccaattcttgattctgcgaagatgcatactgaaatactgacat	1649
QY	1930	ataagggcagacaggaagacagcaggagagactgtgttcttccataaacaatttgactgaat	1989
Db	1650	atgcgcctgcacagcagcgttgcgaacaacctgggcctctgcacacatcttcttaagtaaa	1709
QY	1990	gattgcgaaggttgcctctgcgaattatgaattatcttcggaaagacaaatcagtatcca	2049
Db	1710	aaatagaataataacaaagatttcttgatcttcttcttgatagaagctaaacaagaagtgcct	1769
QY	2050	gaagagactgtatccaatgcgta	2072
Db	1770	tcttggttctgaaataatgctcta	1792

```

RESULT      9
US-09-058-489-14
: Sequence 14, Application US/09058489
: Patent No. 6103886
: GENERAL INFORMATION:
: APPLICANT: Whitehead Institute for Biomedical Research
: APPLICANT: Lahm, Bruce
: APPLICANT: Page, David
: TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
: TITLE OF INVENTION: the Y Chromosome
: FILE REFERENCE: WH197-08PA
: CURRENT APPLICATION NUMBER: US/09/058,489
: CURRENT FILING DATE: 1998-04-10
: EARLIER APPLICATION NUMBER: 60/004,877
: EARLIER FILING DATE: 1997-04-11
: NUMBER OF SEQ ID NOS: 91
: SOFTWARE: fastseq for Windows Version 3.0
: SEQ ID NO 14
: LENGTH: 3408
: TYPE: DNA
: ORGANISM: Human
US-09-058-489-14

```

Query Match	5.7%;	Score 135.4;	DB 3;	Length 3408;
Best Local Similarity	49.2%;	Pred. No. 5.7e-25;		
Matches 454;	Conservative	0;	Mismatches 451;	Indels 18;
				Gaps 3

OY	1165	lctgcltcaaccacccacccggaattagacattccaaagragagaagaaatgttccaaatat	1224
Db	1664	cttgatcttgaccccaacgaagaggttgcgcagctacagatctctcgaaagccgaataattc	1723
OY	1225	tcataataa----ggctcgcagagtglttgbgtacatggtgtgaaataagatatgataca	1281
Db	1724	tcataccgactctagagttctgccttcctgcgtgttctatgtgtgtgcgttatgttgcagc	1783
OY	1282	atagaagagcttaaaaaaagagtgatagatcatcaatacttgcaactcccggaagatgaaatg	1341
Db	1784	atccgagacttggaaacgtgagttgcattgttgatgagcactccagaagctcctagtgat	1843
OY	1342	ctgcgaatgagtaacttcgtccaatcgtgaagatatcaactctgtgtttagatgaaaga	1401
Db	1844	atgatvgaaaagaggaagaatgtgatatgagattcttgcataactctgtgtgtatgataagct	1903
OY	1402	gaaacaaagtcttgcaactgtgatatgtgaaccccgcaataatgaaatgttgttaga-----	1454
Db	1904	gatccgagatcttgatatagtggtttgtgcctcagatctgttagatatgacgaacagatact	1963
OY	1455	-----gtgcgcgcgaagagagcgacagattagaccagttgataatgacctcaatcagtt	1509
Db	1964	atgctcccaaaaggtgtgcgcgcacacatcaatgatagttttagtgcatacttccctaaagaaata	2023
OY	1510	catgcctctgcgaacaactcttatgtgaagaagcaacaatgatactgtctatgttgtatcattg	1569
Db	2024	caagatctgtgcctcgtgattctctatagaatgatatcatcttctgtgcgttgagaagagttgc	2083
OY	1570	ctgattgtgcgtatgattcgaagtgaagaacaaataataatgtgaaaccccgagaagaagaagtg	1629
Db	2084	ctcaactctcgaaa---acatcacacacgaagaagtagtlttgbgtgtgagaagaatcgaacgaag	2140
OY	1630	agtcacaatgcaaaacttcttcaagagtatgtcatccacagaacaagtcatgtcttgcgt	1689
Db	2141	tcattctcgtctgacccctcccaaatgtcaacagagaagaagatctactgaccttagtgttctg	2200
OY	1690	ctccgaagaagctgtgtctgtgagatccactatccaatgtgacccaataactgtgaatatcatgta	1749
Db	2201	gaagacccaanaaaggtgtcagagttctctgcgtgagttctctataccatgaagaatcagcatgt	2260
OY	1750	gagctctctgcatgtgagatagagaacacgagatctcgtggaagaagcatctagaacttaaa	1809
Db	2261	accacagatcccatctggaagacggtctccgaaggtgataagagaagagcccttccacaggtccgc	2320
OY	1810	acagagccaaagtgaaataactaatctgaaactgtgactcgaagcctctcagagactgtatccat	1869
Db	2321	tcaggaanaaaagcccaattttagtgcgtacagagtgacgagcaagaggtactgtgacatttca	2380
OY	1870	gaagtttaacagctgtctataattttttgactttccacaggaatttgaanaatcagttacaacga	1929
Db	2381	aattgtgaaaacatgttatcaaatcttgcactgtgcgaagtgatcatctgaataatgatataactgt	2440
OY	1930	ataaggtcgacaggaagagagagagagatctgtgttctccattcaactctgtactgaat	1989
Db	2441	attgtgtctgaagagacgtgttagaaacctgtgcctgtgcaacccattctttaaagaggg	2500
OY	1990	gatttgagaggtctgcctctgatatgtatataatcttgaaagaagcaaatagatatacttca	2049
Db	2501	aacataaataattactaaagatctgtgtgatactctcttctgtgaagctaaacaaagaagtgcg	2560
OY	2050	gaagaggtctgtacataatgtgactga	2072
Db	2561	ctctgtgttagaanaacatgagctta	2583

RESULT 10-89-13
 US-09-058-489-13
 : Sequence 13, Application US/09058489
 : Patent No. 6103886
 : GENERAL INFORMATION:
 : APPLICANT: Whitehead Institute for Biomedical Research
 : APPLICANT: Lahn, Bruce
 : APPLICANT: Page, David

```

? TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
? TITLE OF INVENTION: the Y Chromosome
? FILE REFERENCE: WH197-08PA
? CURRENT APPLICATION NUMBER: US/09/058,489
? CURRENT FILING DATE: 1998-04-10
? EARLIER APPLICATION NUMBER: 60/041,877
? EARLIER FILING DATE: 1997-04-11
? NUMBER OF SEQ ID NOS: 91
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 13
? LENGTH: 5322
? TYPE: DNA
? ORGANISM: Human
US-09-058-489-13

```

Query Match 5.7%; Score 135.4; DB 3; Length 5322;
Best Local Similarity 49.2%; Pred. No. 6.9e-25;
Matches 454; conservative 0; Mismatches 451; Indels 18; Gaps 3;

QY	1165	ttgattctcaactcccatctcggaagatctgaactccaagttagaaggagaa	tgcttgcaaaat	1224
Db	1664	ttgtgattatgacccaacaaaggagagtgctgcagtaaacagatccacgaagaa	ccagaaaattc	1723
QY	1225	tcataataa---gggcttcggagatcttggtgtatagtggtgtgtgaata	atagaatgaacaa	1281
Db	1724	tcattccgattctagagttcgcttcgcctgcgttggtttatgtgtgtgtgcgaat	atgtgtccagag	1783
QY	1282	atagaagagctctaaanaaagctgtatagatctcaatgtccaactcccggaagt	tttaattgat	1341
Db	1784	attcgagcttggaagcttggaatgcattgttagtcagccccaagcgctcagtgat		1843
QY	1342	ctgcgaatagatcttaacttcgtccaatctgtgaagaataataactcttggttt	atagatgaacaa	1401
Db	1844	atgattgtgaagaagagaaagatttggattttagactcttgcgaataactgtgt	tttagatgaagct	1903
QY	1402	gaacaaagattgttcgacatgtagattttgaacccaagaatagaagatttctga	-----	1454
Db	1904	gactgagatgttgatgtatagggttttgagcccgagttctgtatagaatgtc	tgcaacaagtaact	1963
QY	1455	-----tggtgcgccagatagcgacagcttatgtgaacagtgtacat	tggtcctcaatcagtt	1509
Db	1964	atgcctcccaaaaggtgtccgcgcacactatgtatgttttagtgcactcttc	cttaaggaaata	2023
QY	1510	catcgccctcgacacatcttattcttgaaagaacaaatgatgtctatggttg	taactgtggaatcggtat	1569
Db	2024	cagatgtcgtgtcgatgtattctctagaatgaataatactcttcgtcgtat	gtagaagagtgtgc	2083
QY	1570	ctaatgtctgttaagttcagttggaagcaaaataattgttaaccaacgga	agaagaaatgtg	1629
Db	2084	tcatacctctgnaa--acatcacaaagaaatgatgttttgggtgaagaata	catgacaacaaag	2140
QY	1630	agtcacaaagaaactcttcttaacagagtgtgtcatccacagaacagaat	gtatgtcttcggt	1689
Db	2141	tcattctcgtctgaactcctctaaatgtcaacagcgaagatctcaactga	gtttaggtttgtg	2200
QY	1690	ctctgnaaagctgttcgcgatacactatcaatgtacatgaactgtgaata	ataatacagta	1749
Db	2201	gagacacaaaaggtgtcagatctctctgtgagattctctataccaaatga	agatacagatgt	2260
QY	1750	gagctctcgtacatgtagatgtagaaacaaagagatccggagaaagcatt	gaacttaaa	1809
Db	2261	accagatccatcatgtagagacggtcttcagagggtatagagaagaggtcc	cttaacagttccgcg	2320
QY	1810	acaagcnaagtgtgaatactactaatgtgcaactcgtactcgtctagaag	acttgtatgccat	1869
Db	2321	tcagtaaaaagcccaattttagttgtctacagcatgacgacaaagagatc	gtgacatttca	2380
QY	1870	gacgttaacatgtctataatatttgaatttccacggaatgtgaagaatcgt	taacacaa	1929
Db	2381	aatgtgaacaaatgttataatatttgaactgtccaaagtgtatgtgaaga	aatgtgtactcgt	2440
QY	1930	atagagtcgacaggaagcaggaagatcgtgttctccatbaaactttaga	naat	1989

Db 2441 attgctgcacgagcagcgtgtagaacccttgccctgcgaacctccttaacgagag 2500
OY 1990 gattggaggggtgctcctggaattgattatctcggaaagcgcaatccgagatccca 2049
Db 2501 aacataaataataccaaggaattctggaacctcttgtagaagcacaagaagtgcg 2560
OY 2050 gagagactgtatcaatgcttga 2072
Db 2561 ccttgcttagaacaacatgctcta 2583

RESULT 11

US-09-039-773A-1
Sequence 1, Application US/09039773A
Patent No. 6100388
GENERAL INFORMATION:
APPLICANT: Casas, Ivan
APPLICANT: Jonsson, Hans
APPLICANT: Willestam, Bo
APPLICANT: Roos, Stefan
TITLE OF INVENTION: Lactobacilli Harboring Aggregation and Mucin
Binding Genes As Vaccine Delivery Vehicles
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Standley & Gilcrest
STREET: 495 Metro Place South, Suite 210
CITY: Dublin
STATE: Ohio
COUNTRY: US
ZIP: 43017
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 Inch, 1.44mb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS Version 6.22
SOFTWARE: Microsoft Word Version 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/039,773A
FILING DATE: 16-MAR-1998
CLASSIFICATION: 536
PRIOR APPLICATION DATA: No. 6100388 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Donald O. Nickey
REGISTRATION NUMBER: 29,092
REFERENCE/DOCKET NUMBER: 1229-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 792-5555
TELEFAX: (614) 792-5536
TELEX: No. 6100388 applicable
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Genomic DNA
DESCRIPTION: Genomic DNA sequence and deduced amino
acid sequence of bacterial aggregation
DESCRIPTION: Protein
HYPOTHETICAL: No
ANTI-SENSE: Yes
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Lactobacillus reuteri sp
STRAIN: 1063
CELL TYPE: Unicellular organism
US-09-039-773A-1

Query Match 5.1%; Score 120.6; DB 3; Length 1800;
Best Local Similarity 46.6%; Pred. No. 2.5e-21;
Matches 476; Conservative 0; Mismatches 524; Indels 21; Gaps 2;

OY 971 aaaaataaaaagcgaggtttccaagaacacactatctgaagcagcagcagc 1030
Db 221 AAGCAATCAAGACGAGCGGATACGAAGAAAGCAACCAATTCAAGAAACGATTCCAA 280
OY 1031 ttgtgttcgaagaaatagatctatagagtagagccagacgtggaacaggaagacatcgt 1090
Db 281 TGCTTCTTGAGGGTAAGGATGTATTGTGCACAGACAGACTGGAACTGGTAAGAC----- 335
OY 1091 gtatttaagcctggaattatcatcctcgtgctccttaaccacccagccttaagtcaga 1150
Db 336 -----GCTGCTTTTGGGTGTCCTCAATTATTGAAACGTTGACTGAAATCCCA 385
OY 1151 atagaccgagcatgtagtcttaactcccaactcggaattagcacttcaagtagaaggag 1210
Db 386 ATATTCAAGCAATTATCATTTCAACCAACAGTGAAATTAGCGATTCAGACCACCAAGAAAG 445
OY 1211 aatgttgaataatcatataaaggccttgagtggtttgtatattgtgtgtggaata 1270
Db 446 TTTATCGCTTAGTAAAGATTAACATGTTCCGCTGCAGGTATGTATGGTGGGCGAGATA 505
OY 1271 gagatgaaacaataagaaagccttaaaaaggctgtagatatacataatgcaatcccgaa 1330
Db 506 TTCCGCGCCAAATTAGAGCTTGAAACAAACCCCAAAATTCTCGTGGGACCCCTGGAC 565
OY 1331 gatgaatgacatcgcaaatgagtaacttcgcatactcgaaagaaatataactactgtgt 1390
Db 566 GGTATACGTGACCATATTAAACGTCATACAGTTAAACTTTGACCACTTTAGACCTGTGTC 625
OY 1391 tagatgaaagaaagaaatgttggacatgggaatttgaaacccaataatgaagaatttgt 1450
Db 626 TCGATGAAGCAGATGAAATGCTAAACATGGGATTTCTTAAGAAATTTGAATCATCATCA 685
OY 1451 tagatgtgcgccaagaaagaaagaaatgaaagaaatgaaagaaatgaaagaaatgaa 1510
Db 686 AGGAACACCAAGATGATTCGCAAACTTTCCTTCAGCAACACATCCACCAAGAAATCA 745
OY 1511 atgcgccgcaacaactctatgtgaagaacaaatgattgtctatgttgtatattgattc 1570
Db 746 AGCGAATTTGGGTTCAATTATGCTGATCCGGAACCTGCGGATCAAGGCGCAAGGAAAT 805
OY 1571 tagttgtcgtgaagtctagtgaaagcaaaataataatgttaaccaccgaaagaaatgaa 1630
Db 806 TGACTACTGACTTATGTTGATTCAGTACTA-----TGTTCCGCTCCTGACTATGAAGAT 859
OY 1631 gtcaatcgcaaaactttctcaagagatgtcaatccacagaacaaagtcatgtctcgttt 1690
Db 860 TTGACATCATGACCCGCTTAATTGATGTTCAAGATCCTGATCAATTAATGCTTTGGCTC 919
OY 1691 ctcgaaaagcgtgttcggaatcaactataaagtgaaccttaacttgaatatatacagtag 1750
Db 920 GGACAAAGCGCGGGTAGATGAAATTTGCAAGGGCTTGAATGGCGCTGCGCTACATTCAG 979
OY 1751 agtctcgtcgtgagaaatagaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1810
Db 980 CTGGTATCATGATGATGCTTACTCAGGATGAGCCCTTCTTAAGATCATGATGATTAAGA 1039
OY 1811 caagcaagtgagaaataactaatgtcaacttgaatcagcctctagagagacttgatcagt 1870
Db 1040 ACAATGAACCTTGATATCTTAGTTGCAACAGATGTGGCTGCCGGGCTTAAGCAATTCGG 1099
OY 1871 acgttacaatgtctataatttgacttcaagaaatattgaaagaatcgtacacgaa 1930
Db 1100 GGGTTACGCAATGTTTATATATGATATGATTCATGACCAAGACACTATGTTTACCGGA 1159
OY 1931 taaggcgacgggaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1990
Db 1160 TTGGCCGACACGAGCGCGCGACATCAACGGGATCTTTAATCACTTGTGACTCAATATG 1219
OY 1991 a 1991
Db 1220 A 1220

Db 877 gccacacgagatctctggagatgacacacagttcacgacgacccaatccgacatc---t 933
QY 1557 tggatcattgagatcagttgctgtaagttcagtgaggaataataatgttaaccocga 1616
Db 934 ggtgaaacggtgataatgactcttggaagcatcaagcaaatcttcgtggaagtgaag 993
QY 1617 ggaagagaatlgagtcacatgcaaaatttctacagagatgacttccacagacaagt 1676
Db 994 ggaagagtggaaatttgacactctgtgacctctacgacacactgacactcaactgagc 1053
QY 1677 catgtctcttctcggaaaagctgttcggatataactatcaactgaagtgactataactg 1736
Db 1054 ggtcactctctgcacacaccaaagaagtgacgtgcgcagcagagaagaatlgagggaac 1113
QY 1737 aaatatcagtagagtcctctgcatgagatagagaagagagatccggagagaagcat 1796
Db 1114 caactcactgtlactcctcaatgcatgagacatgcccagaagaagggaggtacatc 1173
QY 1797 agagaaacttaaacagcgaagtggagataactatgtgcaactgactagccctagagg 1856
Db 1174 gaaggaagttccggtcgggcgccagccgagtgctatctctacagatgctcggccagg 1233
QY 1857 actgagttccatgaacgttaccatgctctataatttgaacttccacagaaatgtgaag 1916
Db 1234 gttgagttccctcctaaggtgtccctcactatgaatgacttcccttaataacagagaat 1293
QY 1917 atacgtacaccgaataaggcgccagcaggaagagagactgtgttcttcacta 1972
Db 1294 gtacatacacagaattggggagatcaggtctgataagccggagaaggggtgtgcacat 1349

RESULT 14

US-09-318-443-5
; Sequence 5, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briylanlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIIII, AND METHODS OF USE
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318.443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-443-5

Query Match 4.2%; Score 98.4; DB 4; Length 1536.
Best Local Similarity 47.1%; Pred. No. 1e-15;
Matches 337; Conservative 0; Mismatches 376; Indels 3; Gaps 1;
QY 1257 tgggtgtggaatacagatgaaacaaatagaagagcttaaaaaagtgatataat 1316
Db 496 tggagggaccaaattggtggtgagacatacagaagctgattacgacagcatgtgtcgc 555
QY 1317 tgcaccccggaagattgaatgactgcaaatgaaacttgtaacttgtaactgaagaat 1376
Db 556 gggcaccctcagggcggtgttttgataatgactgcgcgaagaagcctaagacaagtgtcat 615
QY 1377 aaactactggtttatagatgaagcagacagaagtgtgacatgggaatttgaaccagat 1436
Db 616 caaaatgtgtgttcttgatgaaagctgcatgaaatgttgaataaaggttcaaaagacagat 675
QY 1437 aatgaagattctgtagatgtgcccagatagcagacagattataccagatcagtg 1496
Db 676 ttaagatgatataggttaccgtcttcacgacacaggtgtgttcatcagtgccaggt 735
QY 1497 gctcattcagttcattcgccttcgcaaatcttatttgaagaaccaaagatgtctatgt 1556

Db 736 gccacacgagatctctggagatgacacacagttcacgacgacccaatccgacatc---t 792
QY 1557 tggatcattgagatcagttgctgtaagttcagtgaggaataataatgttgaaccocga 1616
Db 793 ggtgaaacggtgataatgactcttggaagcatcaagcaaatcttcgtggaagtgaag 852
QY 1617 ggaagagaatlgagtcacatgcaaaatttctacagagatgacttccacagacaagt 1676
Db 853 ggaagagtggaaatttgacactctgtgtgacctctacgacacactgacactcaactgagc 912
QY 1677 catgtctcttctcggaaaagctgttcggatataactatcaactgaagtgacctaacttgg 1736
Db 913 ggtcactctctgcacacaccaaagaagtgacgtgcgcagcagagaagaatlgagggaac 972
QY 1737 aaatatcagtagagtcctctgcatgagatagagaacagagatctggggagaagcat 1796
Db 973 caactcactgtlactcctcaatgcatgagacatgcccagaagaagcggtgacatcat 1032
QY 1797 agagaactttaaacaagcgaagtggagataactatgtgaactgtactagccctagagg 1856
Db 1033 gaaggaagttccggtcgggcgccagccgagtgctatctctacagatgctcggccagg 1092
QY 1857 actgagttccatgaacgttaccatgctctataatttgaacttccacagaaatgtgaag 1916
Db 1093 gttgagttccctcctaaggtgtccctcactatgaatgacttcccttaataacagagaat 1152
QY 1917 atacgtacaccgaataaggcgccagcaggaagagagagactgtgttcttcacta 1972
Db 1153 gtacatacacagaattggggagatcaggtctgataagccggagaaggggtgtgcacat 1208

RESULT 15

US-09-318-443-1
; Sequence 1, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briylanlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIIII, AND METHODS OF USE
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318.443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Xenopus laevis
US-09-318-443-1

Query Match 4.1%; Score 96.4; DB 4; Length 1245.
Best Local Similarity 46.3%; Pred. No. 3.1e-15;
Matches 479; Conservative 0; Mismatches 531; Indels 24; Gaps 4;
QY 961 gaagtttgaagaacattaaagaagcaggtttcaaaaggcgaacacacattcagttacag 1020
Db 151 gaactcttgagagcatctatgctttagatgtgaaacacatcggcctacacacagaag 210
QY 1021 gcatggccattgtgttgcagaagatagatctatagagtagccagagactgtgaacaga 1080
Db 211 gcaatcaagcatcatcaataaaggagatgtatgcacaaatcacatcagttgtatagcgc 270
QY 1081 aagacatgtgtatttaatagtcgtgagattatcatcactcgtgtcctcaaccacagcttaa 1140
Db 271 aaacagcaacttttggttcgttctgctacaggtgttggatattcagatcgt----- 323
QY 1141 gttcaaaagaaatagaccggcagatgtatgttctaactccactcggaaattagcaactcaa 1200
Db 324 -----tgaaccacagccttgattttagcacccacacaaagattagcagcgcaa 372

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 09:08:45 ; Search time 2465.67 Seconds
(without alignments)
12945.893 Million cell updates/sec

Title: US-09-923-831-42

Perfect score: 2365
Sequence: 1 ttggtaccgagctcgtgctcc.....aaaaaaaaaaaaaaaaaaaaa 2365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gD_estl:*
10: gD_estc:*
11: gD_hic:*
12: gD_gss:*
13: em_gss_hum:*
14: em_gss_luv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	30.1	729	9	AI378017 te67f02.x
2	664	28.1	698	9	AI521173 th58d04.x
3	653.8	27.6	674	10	BG724056 602697033
4	638.6	27.0	742	10	BG772793 602720972
5	622.4	26.3	787	10	BE907039 601499866
6	618.8	26.2	639	10	BG724057 602697035
7	589.4	24.9	599	9	AI637624
8	498.4	21.1	508	9	AI637624
9	448.2	19.0	453	9	AA883800
10	404	17.1	405	9	AA883800 al59d05.s
11	398.2	16.8	403	9	AA948168 cq33b10.s
12	397.8	16.8	404	9	AA948168 hq21h02.x
13	387.2	16.4	404	10	BM341446 hd09d07.x
14	385	16.3	660	9	BM341446 hd09d07.x
15	354.4	15.0	543	10	BE756114
16	354.4	15.0	543	10	BE756114
17	330.6	14.0	679	9	AL637874

18	330.4	14.0	604	10	B7066843
19	311	13.2	574	10	BE895251
20	296.8	12.5	568	10	BG089777
21	285.6	12.1	563	12	BG089777
22	247.6	10.5	562	10	BF367159
23	243.8	10.3	247	9	BE144935
24	238.6	10.1	564	10	B7070471
25	231.4	9.8	809	10	BI464283
26	229.4	9.7	763	12	BH369602
27	218.8	9.3	543	10	BE756114
28	215	9.1	859	10	BI560190
29	210.4	8.6	231	10	BF377109
30	197.2	8.3	406	9	AA781181
31	197	8.3	861	12	CNS06DDO
32	195.8	8.3	980	10	AI518383
33	190.8	8.1	996	12	CNS07228
34	188.8	8.0	902	10	BM453752
35	188.8	8.0	1047	10	BM477615
36	188.8	8.0	1053	10	BM450043
37	188.8	8.0	1103	10	BM464008
38	184	7.8	799	10	BF607995
39	182.6	7.7	825	10	BI250982
40	181.4	7.7	664	9	AV705304
41	181	7.7	801	9	AV136032
42	180.4	7.6	827	12	A2667472
43	177.2	7.5	365	9	AI586361
44	176.2	7.5	735	9	AV126255
45	174.2	7.4	743	9	AV716479

ALIGNMENTS

RESULT 1
LOCUS AI378017/c
DEFINITION te67f02.x1 Soares_NFL_T-GC_S1 Homo sapiens cDNA clone
IMAGE:2091771 3' similar to TR-013370 013370 SUPPRESSOR OF
UNCNTR0LED MITOSIS. ; mRNA sequence.

ACCESSION AI378017
VERSION AI378017.1
KEYWORDS GI:4187870
SOURCE human.
ORANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1048 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 440.

FEATURES

source
1..729
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2091771"
/clone_id="Soares_NFL_T-GC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (pharmacia) with
a modified polylinker; Site_1: Not T; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung, NBH19W, testis NHT, and B-cell
NCL-GAP-CGB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-729111, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. " 2 others

BASE COUNT 196 a 170 c 105 g 256 t

ORIGIN

Query Match 30.1%; Score 711; DB 9; Length 729;
Best Local Similarity 98.4%; Pred. No. 4.4e-97;
Matches 717; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1615 gaggaagaagaatgagtcacatgcacattcttcacagatgcatcaccacagacaa 1674
DB 729 GAAAGAGAGAAATGAGAGTCACATGCAAACTTTCTACAAAGATGTCATCAAGACANA 670
OY 1675 gtcattcttcgcttcgcgaagaagctgtgcgacacattcgaagtcacattctt 1734
DB 669 CTCATTCTCTCTCTTCTTCGAAAAGCTCTGCGGATCCACTTATCAGGTGACCTTATCTCTT 610
OY 1735 ggaataatcatagtagatcctcgcacatgagatagagaacagagagatcggaaga 1794
DB 609 GGAATATATACATAGAGTCTCTGATGAGATAGAGAACAGAGATCGGAGAGACAA 550
OY 1795 tttagagactttaaacagagcaagtgagaaacttaattgcacactgactagacctaga 1854
DB 549 TTAGAGANCTTTAAACAGCGCAAGTGAATACTAATTGSCAACTGATCTAGCCCTAGA 490
OY 1855 ggaactgagtcacatgagcttacacatgctataatttgaacttccacagaaattgaa 1914
DB 489 GGACTTGATGTCATACATGCTTACATCTCATTAATTTGACTTTCACGGAAATTTGAA 430
OY 1915 gaatacgtacacccaagtagagcgacagagagagagagagctgtgttccattaca 1974
DB 429 GAATACGTACACCGAATAGCGCGCACGGGAAGACAGGAGACGTGTTTCCATTACA 370
OY 1975 acctgactgagaatgattgagaggtgtgccttcgaatctgataataatctggaaga 2034
DB 369 ACTTGACTGAGAAATGATTGAGGGTGTGCTCTGAAATGATTAATTTCTGGAAGACGA 310
OY 2035 aatcagaatattccagaagagctgtglatcaatgctgtgagaggtttgagagacatacag 2094
DB 309 AATCAGAGATATCCAGAGAGAGCTGTATCAATGCTGTAGAGGTTTGAAGACATCAACGG 250
OY 2095 aaaaaggaaatggaagaagaatggaagaagctcgaagagagagagagaggttccattaa 2154
DB 249 AAAAGGGAATGGAAGAAATGGAAGAACCTCAAGAGAGGCCCAAGAAAGTTTCATTAA 190
OY 2155 tgccttcgtcactagtcgaggtgtagagaatccaagatttcttaagaatatagtaagaaga 2214
DB 189 TGCTCTGTAAGTACAGTGGGGTACAGATTCATGATTTTAAATATATAGTAAAGAGAA 130
OY 2215 gttattgacatgtgtgacagtatgaaagacccgagctgattgacattgacttaataat 2274
DB 129 GTATTGACATGTTGGCACTATGAAAGACCGGACGATGATTTGATGATTTCTTAATAATAT 70
OY 2275 agtgcctgaaatatagaaatccagtgcttataacttcttataaataagaagattc 2334
DB 69 AGCTTTGAAATATATGAAATCCAGTGTATTACTTCTTATAATAAATATAGATATTT 10
OY 2335 aaacttaa 2343
DB 9 AACTTAA 1

RESULT 2
A1521173/c
LOCUS A1521173 698 bp mRNA linear EST 13-APR-1999
DEFINITION th58d04.x1 NCI CGAP Kid1 Homo sapiens cDNA clone IMAGE:2122471 3'
similar to TR:013370 O13370 SUPPRESSOR OF UNCONTROLLED MITOSIS. ;,
mRNA sequence.
ACCESSION A1521173
VERSION A1521173.1 GI:4435308

KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS 1 (bases 1 to 698)
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdip/image/image.html
Insert Length: 2152 Std Error: 0.00
Seq primer: -400P from G1bco
High quality sequence stop: 424.
Location/Qualifiers

FEATURES
source 1. 698
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2122471"
/clone_1lb="NCI-CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: p7T3D-Pec (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 182 a 158 c 93 g 265 t

ORIGIN

Query Match 28.1%; Score 664; DB 9; Length 698;
Best Local Similarity 97.8%; Pred. No. 4.7e-90;
Matches 673; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 1678 atgcttcctgcttcgaaaagctgtgtgcgatacattatcaagtgcataacttga 1737
DB 698 ATGCTCTCGTTTCAAAAAGCTGTGCGGATCACTAATCAATGACTTAATACGANA 639
OY 1738 aataatacgaatagatctctgcacgtgagaataagaacagagaatcggaagaagatta 1797
DB 638 AATATATCAGTAGAGTCTCTGACGAGATAGAGACACAGAGATCCGGGAAGCATTA 579
OY 1798 gagaactttaaaacaggaagctgagaaataactaattgcaactgactagcctcaga 1857
DB 578 GGAAGCTTTAAACAGCGCAAGGTGAATATCAATTCGAATGATCTAGCGCTTAAGGA 519
OY 1858 ctgattgtccatgacgtttacacatgctcctaatttgcacttccagagaaatctgaaga 1917
DB 518 CTGATGTCACATGACGTACACATGCTATTAATTTTGACTTTCCACGAATATGAGAA 459
OY 1918 taactacacgaatagagcgacaggaagagagagagagagctgtgttccattacaact 1977
DB 458 TACGTACACCGAATAGGGCGACGAGAAAGACGAGAGAGACTGTTCCATTACAACT 399
OY 1978 ttgacagaagaatgattgagaggtgtgctctgaaatgattatcattcgaagaagaagt 2037
DB 398 TTGACTAGAAATGATGAGGGTGGCTCTGATGATTAATATTTCTGAAAGACCAAT 339
OY 2038 cagagatccagagagctgtglatcaatgctgagagaggtttgagacatacaacgaa 2097

```
|||||
Db 338 CAGAGTATTCAGAGAGCTGTGTCAATGCTGAGAGGTTTAAGGCACATCAGACAGAA 279
QY 2098 agggaaatgaaagaaatgaaagaccccaaggaagcccaagaattcattatgtc 2157
Db 278 AGGGAATGGAAGAAAAATGGAAGACCTCAAGAGAGCCCAAGAACTTTCATTAAATGT 219
QY 2158 ctctctactagctgggttagaagaattctttagaataagtagaagacagaagta 2217
Db 218 CTTCTGTACTAGTGGGGTAGAGAAATTCAGATTTTATAGAAATATAGTAAGCGAGAGTA 159
QY 2218 ttgagacatgtgacagatgaaagacccgacgactgattacgattcttaataaataag 2277
Db 158 TTGGACATGTTGGCAGTATGAAGAGACCGACATGATTGCTTAATAATATAGT 99
QY 2278 gtttgaaatataagaatcagtgcttatactctcttaataaataagaagtaattaa 2337
Db 98 GTTGAATAATATAGAAATCCAGCTTTTATCTTTTAAATAATATAGATTTTAA 39
QY 2338 cttaaaaaaaaaaaaaaaaaaaaaa 2365
Db 38 CTTGGAATAATAAGAAAAA 11

RESULT 3
LOCUS BG724056 674 bp mRNA linear EST 08-MAY-2001
DEFINITION 60269703F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4829072 5',
ACCESSION BG724056
VERSION BG724056.1 GI:14003243
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 674)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10748 row: a column: 09
High quality sequence stop: 674.
Location/Qualifiers
FEATURES
source
1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4829072"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 219 a 137 c 182 g 136 t
ORIGIN

Query Match 27.6% Score 653.8; DB 10; Length 674;

Best Local Similarity 99.6%; Pred. No. 1.6e-88;
Matches 666; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 183 gtcgagcgcgcctctcttggaacaaatgtccacacagagagctcccaagccttac 242
Db 6 GCGGAGACCGCCCTCTTGTGGAACATGTCCACACGAGAGACTCCAAAGCCCTTAC 65
QY 243 gtgggtctgtctgtctgctgagcagctgagctgctgctgctgctgctgctgctgct 302
Db 66 GTGGGTCTTCTGTCTGTGTGCGGAAAGCTGCACTGTCCGAGCCGACAGAGAGGCGGCG 125
QY 303 ggaagagltgaatgcgaaggtctggaaggaatagtgctgcgcagagtgctgcctgag 362
Db 126 GGAGAGATGTGAATGGAACAGTCTGAGGATATAGTGTGCGCAGAGTGTGTGCGAG 185
QY 363 aggcacccctagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 422
Db 186 AGGCACCTCTAGGCGCCCGGAGAGACGTGGCGCTGGTCAGAGAACTGCCCTGTGTTT 245
QY 423 tgcttgaagagcacttgctgagcgtgtaatcgtctgctgctgctgctgctgctgctgct 482
Db 246 TGCTTTGAAGAGCCACTTTGTTGGCGGTATGCGTGTGCGTGTGCGTGTGCGTGTGCGT 305
QY 483 tatcaaatgatacaacaacacacacacacacacacacacacacacacacacacacacac 542
Db 306 TATCAAAAGTACAAACAAACCAACCAATTCATTAATACAAACCAAGAAATCAATTAAT 365
QY 543 caaaatcttggaagcagagcaatgcgaacgaagcaaaagcagtgatagacatctgtc 602
Db 366 CAAAATTTTGGCGAGCAGAGCAATGCAAAAGCAAAAGCAAGTATGACAAATTTTGT 425
QY 603 taagaagctgaagaagaattacacatcagaatgcgaatgatactgcatccacccctc 662
Db 426 TAAAAGCTGAGAGAAATTAATACAAATTCAGAAATGCGGAATGATCTGCAATCAACCTTC 485
QY 663 tgttggaagcagagcagacagatacaaatgtctggaagagatcgccatctgtag 721
Db 486 TGTGGAAGAGATGAGAACACAGATTAACATGTTGTCGAGAGATCCGCATGTAGTAG 545
QY 722 attggatcaaatgaagagaggtctggaatgcaaaaacaaagtgagagattac 781
Db 546 ATTGGATCAAAATTAGAGAGAAAGTTTGAATATGCAAAAACAAAGTGGAGATTTTAC 605
QY 782 caccacataagaagaacatttataagaagctcactgcacaaagtgcattcaaaagtag 841
Db 606 CACCAATTAAGAAAACTTTTATTAAGAGTCCACTGCCACAGTGCATGTCAAAAGTAG 665
QY 842 aagcagata 850
Db 666 AAGCAGATA 674

RESULT 4
LOCUS BG72793 742 bp mRNA linear EST 15-MAY-2001
DEFINITION 602720972F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4837888 5',
ACCESSION BG72793
VERSION BG72793.1 GI:14083446
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 742)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
```

cdna Library Arrayed by: The I.M.A.G.E.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E.E. Consortium/LNL at:


```
QY 1854 aggaacttgatgltccatgaacgttacacatgltcataattcttgacttccacgaatatga 1913
|||||
Db 301 AGGACTTGATGTCATGACGTTCACATGTCATATATTTCACCTTCCAGGAAATTTGA 360
|||||
QY 1914 aaaaatcgttacacccgaatagggcgacgggaagacgaaggagactggtgttccattac 1973
|||||
Db 361 AGAATACGTACACCGAATAGGGCGACGGGAAGACAGGAGGAGCTGTGTTCATTCAC 420
|||||
QY 1974 aacttgactagaatgatgtggaagggttcgctcgtgaattgataatacttcggaagagc 2033
|||||
Db 421 AACTTGAAGTGAATGATTTGAGGGGTGCTCTGATTTGATTTATTTCTGGAAGAGAC 480
|||||
QY 2034 aatacagaagtatcccaagaagacttgatcaatgagctgaagggttcgaagacataacg 2093
|||||
Db 481 AATATGAGATATTCACAGAGAGCTTTATCAATGAGCTGAGAGGTTTGAAGCATTCACG 540
|||||
QY 2094 gaaaaggg---aaatggaagaaatgaaagaccccaaggagagcccaagaatttc 2149
|||||
Db 541 GAACACAGGACAAATGGAACCGAATAATGGAAGACTCAAGAGAGGCCACAGAG -TTC 599
|||||
QY 2150 attaatgctctctgactagtggtgtagag---aattcaagaatttttagaataata -gt 2205
|||||
Db 600 ATTAAATGCTCTGCTACTAGTGGGCTAGAGAAATTCACGATTTTAAATAATATAGGT 659
|||||
QY 2206 aagacag-aagtalctgacatgtgtgcaatgaa--gagacggactgacttgactgat 2262
|||||
Db 660 AAGACAGAAAGTATCGACATGTTGGCCGCTACTGACGAGGCCGAGCGTTTACCTGAT 719
|||||
QY 2263 tcttaaaataatagtggttgaataatagaacacagtggtttaaacttcttcaataaa 2322
|||||
Db 720 CCTTCACACTTATAGTTCG-AACCTTTGAATCCAGGGTGTATATCTTCTTCATCAAAA 778
|||||
QY 2323 atag 2326
||
Db 779 TGAG 782

RESULT 6
Bg724057 639 bp mRNA linear EST 08-MAY-2001
LOCUS 602697035F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4829168 5',
DEFINITION mRNA sequence.
ACCESSION Bg724057
VERSION Bg724057.1 GI:14003244
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 639)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue procurement: Miklos Palokovics, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10748 row: e column: 09
High quality sequence stop: 636.
Location/Qualifiers
1..639
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4829168"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"

FEATURES
Source

Query Match 26.2%; Score 618.8; DB 10; Length 639;
Best Local Similarity 99.5%; Pred. No. 2,7e-83;
Matches 631; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 183 gtcgagcgccccctcttggaacaatgtccaccacggaagagctccaaagcctcac 242
|||||
Db 6 GCGGAGCGCCCTCTTGTGAAATGTCCACACGAGGAGCTCCCAAGGCCCTCTAC 65
|||||
QY 243 gtcggtcgtctgactgagcgaagctcgacagtgctccgagcgcaagagagcgagc 302
|||||
Db 66 GTGGGTGCTGCTAGTGTGCGAGAGCTGCAGAGTGTCCCGAGCGCCAGAGAGGCGCGC 125
|||||
QY 303 ggaagagttgaatcgacacaggtcctgagagataatggttcgcgacagagtggtcgtgag 362
|||||
Db 126 GGAGAGTTGAATTCAGACAGGTCTGAGAGGATATAGTGTGCGACAGAGGTGTCTGTGAG 185
|||||
QY 363 aggcaccttaggcccccgagcggtgagcgctggtcagcagaggaactgcgctgttt 422
|||||
Db 186 AGGCACCTTAGGCCCCCGAGAGAGTGTGCGCGGTGTCAGAGAACTGCGCTGTGT 245
|||||
QY 423 tgccttgaagagccacttggtgagcggtgaatcgtgtgtgtgttcaaaaataagaa 482
|||||
Db 246 TGCTTTGAGAGAGCCACTTTGTTGGCGGTAAATGCGTGTGCGTCAAAAATPAAGAA 305
|||||
QY 483 tatacaagtacaaacaaacacccaatccaaataataatacaagaacaacagaatcattagt 542
|||||
Db 306 TATACAAAGTACAAACAAACACACACATCCAAATPAATACAAACCAACCAATCATTAGT 365
|||||
QY 543 caaatcttggcagcgaagcgaatgcaaacgaaagaaagagatagacaaatttgt 602
|||||
Db 366 CAAATTTTTCGACAGCAAGCAATGCAAAAGCAAAAGCAAGATGATAGCAATTTTGT 425
|||||
QY 603 taaaagctagaagaataatcaaatcagaatgcggaattgatactgcattccaacttc 662
|||||
Db 426 TAAAAAGCTAGAAAGAAATTAACAATTCAGATCCGAATTTGATCTGATTCACAACTTC 485
|||||
QY 663 tgttgaaagaatgtgaagacacagatacaatgtgtgtgacagagatgagcattgata 722
|||||
Db 486 TGTTGGAAAGATGAGACACAGATMACATGTTGTTGCAAGATTCGGCCATTTGATAGA 545
|||||
QY 723 ttgggatcaaatatgagagagaggtttgaaatggcaaaaacaaagtggcgagatttacc 782
|||||
Db 546 TTGGGATCAAAATTAAGAGAGAGGTTTGAATAG-AAAAAACAAGTGGCGAGATTAC 604
|||||
QY 783 accaattagaagaaactttataaagagttcact 816
|||||
Db 605 ACCAATTAGAAAGAACTTTATTAAGAGTCCACT 638
|||||

RESULT 7
Ai637624 599 bp mRNA linear EST 14-DEC-1999
LOCUS tli0c11.x1 NCI_CGAB_G6 Homo sapiens cDNA clone IMAGE:2240372 3',
DEFINITION similar to TR:059857 059857 MOC2 RNA HELICASE. [1] ;, mRNA
sequence.
ACCESSION Ai637624
VERSION Ai637624.1 GI:4689858
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 599)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLML at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 495 Std Error: 0.00
Seq primer: -400P from Glbco
High quality sequence stop: 444.
Location/Qualifiers

FEATURES
source

1..599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2240372"
/clone_lib="NCI-CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; plasmid DNA
from the normalized library NCI-CGAP_GC4 was prepared, and
as circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 159 a 141 c 78 g 221 t
ORIGIN

Query Match 24.9% Score 589.4; DB 9; Length 599;
Best Local Similarity 99.0%; Pred. No. 6,9e-79;
Matches 593; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1733 tctctgcatgagatagagaaacagagatcgggaaagacattagaaccttaaacaa 1812
|||||
DB 599 TCTCTCATGAGATAGAGAACAGAGATCGGAGAAACATTAGCAACTTTTAAACA 540
QY 1813 ggcgaagtgagaaactaaatctgaactgactctagaggaactgattgcatgac 1872
|||||
DB 539 GGCAAAGTGAGAACTACTAATTCGCAACTGATTCACCTCTGAGAGACTTGATCCATGAC 480
QY 1873 gttacacatgctctataatttgaacttccacggaatactgaagaatgctacacggaata 1932
|||||
DB 479 GTTACACATGCTCTAATAATTTTGAATTTCCACGGAATTTTGAATAATCTACACCGAATA 420
QY 1933 gggcgacgggaaagcagggaggaactggttctccaattcaacttgcactgaataat 1992
|||||
DB 419 GGGCCGACGGAGAGACGAGGACGCTGTTCCATTTCACACTTTGACTGGAATGAT 360
QY 1993 tggagggctgcctctgaattgaattaatatcttgaaaagcaaatcagaagtatctcagag 2052
|||||
DB 359 TGGAGGGTGTCTCTCAATTTGATTAATTTCTGGAAAGACCAATATGAGATTTTCCAGAG 300
QY 2053 gagcttctatcaatgctcagaggttctgagcacatcaacggaagaaggaatctgaaga 2112
|||||
DB 299 GAGCTGTATCATATGCTGACGAGGTTTAAGCAGACATCAACAGAAAGGGAATGGAAGA 240
QY 2113 aaaaatgaaagacctcaaggaagcccaagaagttcattcaatgcttcttactagtgg 2172
|||||
DB 239 AAAATGGAAGACCTCAAGGAAGGCCCAAGAAATTTCATTAAATGCTCTGTACATCAATGGG 180

QY 2173 ggtagagaattcaagattttttgaataatagtaacagaagatattgagcatgttgcca 2232
|||||
DB 179 GGTAGAGATTTCAAGATTTTTTTTGGAAATTTAGTAAGACGAGATTTGCACATGTTGGCA 120
QY 2233 gattgaagagaccggaactgattgactgattcttaataatagtgcttgaaataataga 2292
|||||
DB 119 GATGAAAGACCGCGACTGATTTGAAATGATTTTAAATATGTTGAAAAATATAGA 60
QY 2293 atccagtgcttacttcttcttaataaataagagatatttaacttaaaaaa 2351
|||||
DB 59 ATCCAGTGTATTACTTTTCTTTAATAAATAATAGAGATTTTAAATGAAAAA 1

RESULT 8
AL043701/c 508 bp mRNA linear EST 29-FEB-2000
LOCUS
DEFINITION DKF2p34K0927.s1.434 (synonym: htes3) Homo sapiens cDNA clone
DKF2p34K0927.3', mRNA sequence.
ACCESSION AL043701
VERSION AL043701.1 GI:5423087
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 508)
Blecker, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.

REFERENCE
AUTHORS
TITLE EST (Blecker, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Blecker H

MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No RI sequence available.
This clone (DKF2p34K0927) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers

FEATURES
source

1..508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p34K0927"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 143 a 117 c 71 g 177 t
ORIGIN

Query Match 21.1% Score 498.4; DB 9; Length 508;
Best Local Similarity 98.8%; Pred. No. 2,9e-65;
Matches 502; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1824 aatacctaattgcaattgctcagctctgaggaactgattgcatagacttacaatgt 1883
|||||
DB 508 AATACTAATTTGCACTGATCTAGCCTCTTACAGCACTTGATGTCATGACGTTACATATGT 449
QY 1884 ctataatttgacttccacggaatctgaagaatcgtacacacgaatagggcgcaagg 1943
|||||
DB 448 CTATAATTTTGACTTTCCACGGAATTAATTAATATCAACGCAATAGGCGCTCGCG 389
QY 1944 aagaagcagggaggaactggtgttccattacaacttggactgaagaatattggagggcttcg 2003
|||||
DB 388 AAGAGCAGGAGGAGCTGCTGTTTCCATTACACTTTGACTAGAAATTTGAGAGGTTGC 329

```
QY 2004 cctcgaattgataataatcttcggaagaagacaatcagaatattccagaaggattgatac 2063
|||||
Db 328 CTCGATTTGATTTAATTTCTCGAAGACCAATACAGCTTTTCACAGAGAGCTTTATTC 269
QY 2064 aatgctcgaagggtttgagcacatacgaagaaaggaaatgtgaaagaaatgtgaaag 2123
|||||
Db 268 AATGGCTGAGAGGTTTAAAGCACATCACAGAAAGGCAAAATGCAAGAAATGGAAG 209
QY 2124 accctagaagaagcccaagaagttcattcaatgctcttcgtcagtcagtgaggatgaaatc 2183
|||||
Db 208 ACCTCAGAGAGGCCCAAGAAAGTTTCATTATGCTTCTGTACAGTGGGCTAGACAATTT 149
QY 2184 caagaatttttaagaataatagtaagacaagaatattgacatgltggcagatgaagaga 2243
|||||
Db 148 CAAGATTTTGTAGAAATATAGTAAAGACGGAAGTATTTGGACATGTTGGCAGTATGAAGAGA 89
QY 2244 ccgagactgattgactgattccttaataatagtggtttgaaatatagaaatccagtggtt 2303
|||||
Db 86 CCGAGCTGATTTGACGTATCTTAAATAATAGTGTGTTGAAATAATATGAAATCCAGTGT 29
QY 2304 tatacttcttctaataaataagaagta 2331
|||||
Db 28 TATACTTCTTTAATAAATAATGAGTA 1

RESULT 9
AA83800 453 bp mRNA linear EST 04-JAN-1999
LOCUS al59ad05.sl Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1461609 3' similar to SW:DB10_NICSY P46942 RNA HELICASE-LIKE
PROTEIN DB10. ;, mRNA sequence.
ACCESSION AA83800
VERSION AA83800.1 GI:2993330
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (Info@image.llnl.gov) for further information.
Possible reversed clone; similarity on wrong strand
Insert length: 855 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham
High quality sequence stop: 280.
Location/Qualifiers
1..453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1461609"
/clone_11b="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHR, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 125 a 110 c 58 g 160 t
ORIGIN
```

```
Query Match 19.0%; Score 448.2; DB 9; Length 453;
Best Local Similarity 99.3%; Pred. No. 9,6e-58;
Matches 450; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1891 ttgacttccagaagataatgaagaatagtaacccgaatagagccaggaagaga 1950
|||||
Db 453 TTTGACTTCGACAGGAATTTGAGCAATACCTACACCGAATAGCGCGACGGAGAGCA 394
QY 1951 gggagagctggtgttccattacaaacttgactagaatgattgaggggttgccctgaa 2010
|||||
Db 393 GGGAGGAGCTGCTGTTCCATTCAACTTGTGCAATGATGAGGGGCTGCCCTGAA 334
QY 2011 ttgataataatctcgaagaagacaatcagaatgataccagaagcttgataatgct 2070
|||||
Db 333 TTGATTAATTTCTGGAAGACCAATACAGATTTCACAGAGAGCTTGTATCAATGGCT 274
QY 2071 gagaagcttgaggcacatacgaagaaaggaaatgtgaaagaaatgtgaaagacctcaa 2130
|||||
Db 273 GAGAGGTTTGAAGCACATCACGAAAGGCAAAATGCAAAATGGAACGAGCTCAA 214
QY 2131 ggaaggcccaagaagttcattcaatgctcttcgtcagtgagggtagaagaatcaagatc 2190
|||||
Db 213 GGAAGGCCCAAGAGTTTCATTAAATGCTTCTGTACAGTGGGCTAGAGAAATTCAGATT 154
QY 2191 tttaagaataatagtaagacagaagatttgacatgttgacagatgaagaccggact 2250
|||||
Db 153 TTTTGAATAATATAGTAAAGACAGAAATATTTGGACATGTTGGCAGTATGAAGAGACCGACT 94
QY 2251 gatttgactgattccttaataatagtgcttgtaaaatatagaatccagtgatttactt 2310
|||||
Db 93 GATTTCAGTATCTTAAATAATATATGTTGAAATAATATGAAATCCAGTGTATATCTT 34
QY 2311 tctttaataaataagaagtttaacttaa 2343
|||||
Db 33 TCTTTAATAAATAATGAGATTTTAAACTTAA 1

RESULT 10
AA948168 405 bp mRNA linear EST 23-JUL-1998
LOCUS AQ33b10.sl NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1588123 3',
DEFINITION mRNA sequence.
ACCESSION AA948168
VERSION AA948168.1 GI:3109421
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.llnl.gov/bdrp/image/image.html
Insert length: 484 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham.
Location/Qualifiers
1..405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1588123"
```

/clone.lib="NCI CGAP GC4"
/tissue.type="pooled germ cell tumors"
/lab.host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 100 c 48 g 143 t
ORIGIN

Query Match 17.1%; Score 404; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 4e-51;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1937 gcaaggaagagcagagagagctggtgttccattacaacttgactagaatgttga 1996
DB 405 GCACGGAGAGACGAGGAGCTGTGCTTCAATTCACACTTGTGACTAGAAATGATTCGA 346
OY 1997 gggctgcctcgaattgataataatctcgaaagagcaatccagatccagaagagc 2056
DB 345 GGGTGGCCTCGAATTGATTATTTCTGAAAGACGAATTCAGACTATTCAGAGAGAC 286
OY 2057 ttgtatcaatgctcgaaggttcttgagcacatcaacgaaagaaatgaaagaaaaa 2116
DB 285 TTGTATCAATGGCTGAGAGGTTTGAGSCATCAACGAAAGGAAATGGAAGAAAAA 226
OY 2117 tggaaagacctcaaggaagcccaagaagttcattatgcttcctgactagtg99tla 2176
DB 225 TGGAAAGACCTCAAGAGAGGCCCAAGATTTCATTAATGCTTCTGTACTAGTGGGTA 166
OY 2177 gagaattcaagatttttagaataatagtagacagaagatggacatgttgagat 2236
DB 165 GAGATTCAAGATTTTTGAATAATAGTAAGACAGAAATGGAATGGCAAGAT 106
OY 2237 gaagagaccggaactgatttgactgattcttaaaaaaaatagtglttgaataatagaatcc 2296
DB 105 GAAGAGACCGGACTGATTGACTGATCTTAATAATTAATAGTGTGAAAAATATGAAATCC 46
OY 2297 agtgtttatacttctttaaataaataaagaatattaaact 2340
DB 45 AGTGTTTATACCTTCTTTAATAATAAATAGATATTAACTT 2

RESULT 11 405 bp mRNA linear EST 22-MAR-2000
AM593728/c hg21h02.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2946291 3'
LOCUS similar to TR:015919 015919 MITOCHONDRIAL DEAD BOX PROTEIN.; mRNA
DEFINITION sequence.

ACCESSION AM593728
VERSION AM593728.1 GI:7280986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 405)
TITLE NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/resources.shtml
Seq primer: -400P from Gibco.

FEATURES
source location/Qualifiers
1. 405

/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:2946291"
/clone.lib="NCI CGAP GC6"
/tissue.type="pooled germ cell tumors"
/lab.host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 100 c 48 g 143 t
ORIGIN

Query Match 17.1%; Score 404; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 4e-51;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1937 gcaaggaagagcagagagagctggtgttccattacaacttgactagaatgttga 1996
DB 405 GCACGGAGAGACGAGGAGCTGTGCTTCAATTCACACTTGTGCTAGAAATGATTCGA 346
OY 1997 gggctgcctcgaattgataataatctcgaaagagcaatccagatccagaagagc 2056
DB 345 GGGTGGCCTCGAATTGATTATTTCTGAAAGACGAATTCAGACTATTCAGAGAGAC 286
OY 2057 ttgtatcaatgctcgaaggttcttgagcacatcaacgaaagaaatgaaagaaaaa 2116
DB 285 TTGTATCAATGGCTGAGAGGTTTGAGSCATCAACGAAAGGAAATGGAAGAAAAA 226
OY 2117 tggaaagacctcaaggaagcccaagaagttcattatgcttcctgactagtg99tla 2176
DB 225 TGGAAAGACCTCAAGAGAGGCCCAAGATTTCATTAATGCTTCTGTACTAGTGGGTA 166
OY 2177 gagaattcaagatttttagaataatagtagacagaagatggacatgttgagat 2236
DB 165 GAGATTCAAGATTTTTGAATAATAGTAAGACAGAAATGGAATGGCAAGAT 106
OY 2237 gaagagaccggaactgatttgactgattcttaaaaaaaatagtglttgaataatagaatcc 2296
DB 105 GAAGAGACCGGACTGATTGACTGATCTTAATAATTAATAGTGTGAAAAATATGAAATCC 46
OY 2297 agtgtttatacttctttaaataaataaagaatattaaact 2340
DB 45 AGTGTTTATACCTTCTTTAATAATAAATAGATATTAACTT 2

RESULT 12 403 bp mRNA linear EST 31-JAN-2000
AM341446/c hd09d07.x1 Soares_NFL_T_GBC.S1 Homo sapiens cDNA clone
LOCUS IMAGE:2909005 3' similar to SW:DBP3_YEAST P20447 PROBABLE
DEFINITION ATP-DEPENDENT RNA HELICASE DBP3.; mRNA sequence.

ACCESSION AM341446
VERSION AM341446.1 GI:6838072
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 403)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400P from Gibco.

FEATURES
 source
 1. 403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2909005"
 /clone_lib="Soares_NFL_T-GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: p77T3D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CGAP GC81) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 114 a 97 c 46 g 146 t
ORIGIN

Query Match 16.8%; Score 398.2; DB 9; Length 403;
Best Local Similarity 99.3%; Pred. No.: 2.9e-50;
Matches 400; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1940 cgggaagaacgagagagcgtgcttccatcacacttgcacgaatcagaatcagatgaggg 1999
 |||||||
 Db 403 CGGGAAGACGAGGAGGAGCTGGCTTTCCCTTACAACTTTGACAGAAATCATTTGAGGG 344
 |||||||
 Oy 2000 ttgctctgaattgatttaattctcgaaagacaatcagaatcctcagagagcttg 2059
 |||||||
 Db 343 TTGCTCTGCAATTTGATTAATCTGGAAGACCAATTCAGAGTTTCCAGAGAGCTTG 284
 |||||||
 Oy 2060 tatcaatgagctgaagaggtttgagggcacatcaacgagaaagagaaatgaaagaaatg 2119
 |||||||
 Db 283 TATCAATGGCTGAGAGGTTTGAAGCAGATCAACGAAAGCAATGSAAGAAATGCG 224
 |||||||
 Oy 2120 aaagacctcaagaaagcccaagaagttcaatcaatgctctcgtactagtggggtagag 2179
 |||||||
 Db 223 AAAGACCTCAAGAAAGCCCAAGAGTTTCATTAAATGTTCTGTACTAGTGGGCTAGAG 164
 |||||||
 Oy 2180 aatccaagaatttttagaataatagtaagacagaagtattgacatggttcagatagaa 2239
 |||||||
 Db 163 AATTCAAGATTTTATAGAAATATAGTAAAGACAAATATATGACATGTTGSCATATGAA 104
 |||||||
 Oy 2240 gagaacgagactgattgactgaattcttaaaataagtggttgaataatagaatccagt 2299
 |||||||
 Db 103 GAGACCGAGCTGATTTGACTGATTTTAAATATATGTTTGAATAATATGAAATCCAGT 44
 |||||||
 Oy 2300 gtttatacttcttaataaataagaatatttaactaa 2342
 |||||||
 Db 43 GTTTTACTTCTTTATATAAATATGAAATATTTAAATTA 1

RESULT 13
 BF193072 564 bp mRNA linear EST 02-NOV-2000
LOCUS BF193072
DEFINITION 244514 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF193072
VERSION BF193072.1 GI:11076441
KEYWORDS EST.
SOURCE pig.

ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE
 1 (bases 1 to 564)
 Fahnenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grosse, W.W., Bennett, G.A., Laegreid, W.W.,
 and Keele, J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith RPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

PCR primers
 FORWARD: AGGAACGCTATGACCAT
 BACKWARD: GTTTCCAGTCAGACG
 Plate: 71 Row: L Column: 19
 Seq primer: ATTAGTGACACTATG.

FEATURES
 source
 1. 564
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORF6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 180 a 103 c 129 g 152 t
ORIGIN

Query Match 16.8%; Score 397.8; DB 10; Length 564;
Best Local Similarity 82.4%; Pred. No.: 3e-50;
Matches 472; Conservative 0; Mismatches 92; Indels 9; Gaps 1;

Oy 901 gagaagacactatcccaatccatccgcagcatltagagcgccttcaatgatact 960
 |||||||
 Db 1 GAGAAAGCTCTATATCCCAATCCCTGTTAAATTTGAGGATTCAGTACGTTACCT 60
 |||||||
 Oy 961 gaggtttagaataaataaagcaggttttcaaaagccaaactattcaagcaag 1020
 |||||||
 Db 61 GAAGTATGATTAACATTAAAGCGAGGTTTCAAAAGCCAAACCAATTCAGTCACAG 120
 |||||||
 Oy 1021 gcatggccatctgcttgcagaagatagatctctaaagtagagtaagccagactggaacagga 1080
 |||||||
 Db 121 GCATGGCCAAATCATTTACAAAGAAATGATCTTTATAGAGTAGCCGACGACTGAAACAGG 180
 |||||||
 Oy 1081 aagaatctgttatcttaagctcgtgatcttcaatctcgtctcctcaaccgccttaa 1140
 |||||||
 Db 181 AAGACATTGTCTTACCTAATGCTGGATTTATCTAGTTGACTCACAAAC-----T 231
 |||||||
 Oy 1141 ggtcaaggaatagaccggcagatgagttcaactcccaactcgggaatlagactcaa 1200
 |||||||
 Db 232 GTAGCCAGCATGGCCCGGCGCATGTAGTCTTACACCACCACTGAGAAATTAGCTGCA 291
 |||||||
 Oy 1201 gtaagaagagatgtgcaataatcatataaaggcctcgggggtgttgtatagtt 1260
 |||||||
 Db 292 GTGGAAGCTGAGTGTCTTAAGTATTCATTAAGGCTTTAAAGGTGTGTGATATGGT 351
 |||||||
 Oy 1261 ggtggaatlagagatgaacaatagaagacttaaaaaaggtgtagatatactaa 1320
 |||||||
 Db 352 CGTGAGATCGAGATGACAAATAAAGACTTATTAAGAGGTGATATCATCATTCGA 411
 |||||||
 Oy 1321 actccggaagaatgaataatctgcaaataggatgaacttcgaactggaagaataaac 1380
 |||||||
 Db 412 ACTCTGGAGGCTCAATGATCTGCAAATGAATACCTTTGTGTACTTGAAGACATTAAC 471

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES
source

Location/Qualifiers
1..660
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5930424019"
/clone_1db="RIKEN full-length enriched, 13 days embryo forelimb"
/sex="mixed"
/tissue_type="forelimb"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site_1: Salt: Site_2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGATTAATTAATATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from lambda PLC I. Cloning sites, 5' end: Salt: 3' end: BamHI."

BASE COUNT 228 a 109 c 150 g 173 t
ORIGIN

Query Match 16.3%; Score 385; DB 9; Length 660;
Best Local Similarity 77.1%; Pred. No. 2.3e-48;
Matches 513; Conservative 0; Mismatches 135; Indels 17; Gaps 3;

Qy 1659 gtcatccacagcaagtcattgctcgt-ttcgcgaagctgttcggtacattat 1717
Db 1 gtcacctaaagacaaagtcattgtttgtcagcagaaaaagccgttcgtacattat 60
Qy 1718 caagtgaccttaacttggaaatatcaagtagagctctgcattgagatagagaacaga 1777
Db 61 caagtgatctgattctccgacacatcagtgagagctctgcacgcaacagacagacaga 120
Qy 1778 gaagtcgaggaagaacattagaacatttaaacagcaagtgagaataactaatgcaa 1837
Db 121 gtaaccgagaaagcgttaagaaatttaagacaggttaagtgagaattcttattgtcta 180
Qy 1838 ctgacttaagcctctagagagacttgatgtccatgacgtttacacatgtctataatttga 1897
Db 181 ctgacttgagcattctgagagctctgagagctctgagagctctgagagctctgagag 240
Qy 1898 ttccaggaatatgaagaatactacacaggaatagagcgacggaagaagcaagagaga 1957
Db 241 ttccaggaacattgaaagaaatgacacagagagcgacacacgagagcagagcgacga 300
Qy 1958 ctggtgttcccaacttgacacttgacgaatgacgtgagaggttgctcttgaaattgata 2017
Db 301 ctggaattgtcaatttaccctttacacaaagaatgatggagcggtgccactgaattgattga 360
Qy 2018 atattctggaagaagaacaaatcagagatttcagagagagcttgatcaaatgctgagaggt 2077
Db 361 atattctgaaagacaaatcagaattatccagagagaaacttgcttaattgctgagagatt 420

Qy 2078 ttgaagcacatcaacggaaggaaggaatgaaagaagaaatgaaagacctcaaggaaggc 2137
Db 421 acaaagcaaatataacttaaaagagagatgcaaaaaaaatggcaagacctcaagcaaac 480
Qy 2138 ccaagaagttcatttaactgtctctgtactagtgaggtgagaaattcaagattttaga 2197
Db 481 ccacagaagtgttactaatgtttgcgtgagacactagtcgccacagagatttcaagattgtagt 540
Qy 2198 aatatagtaagacagaaatgtgacatgttgcaatataagaagaccggaactgattga 2257
Db 541 caaagtg-----gaagctgttaaacattacgtcactatgagagagac-----ttgg 584
Qy 2258 ctgattcttaaatatagtggttgaaataatagaatccagtggtttatacttcttaa 2317
Db 585 ctgatttcttaaaatagtggttgaatgtatgaattcaattgtttattcttttaa 644
Qy 2318 taaga 2322
Db 645 taaaa 649

Search completed: July 25, 2002, 11:05:24
Job time: 6999 sec

30-OCT-1998; 98US-0183789.

Human lung tumour
Human C13J3 prostate
T. thermophilus HcL
Staphylococcus aure
Staphylococcus aure
Drosophila melanog
Drosophila melanog
Drosophila melanog
Staphylococcus aure
S. aureus DNA encod
S. aureus dppA cod
Human immune/thema
Human immune/thema
Human immune/thema
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
Human immune/thema
Staphylococcus aur
Human vasa cDNA
Arabidopsis thaliana
Arabidopsis thaliana
Drosophila melanog
Human breast cancer
Human breast cancer
Bacillus anthracis
Enterococcus faeca
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Drosophila melanog
Arabidopsis thaliana
Drosophila melanog
Dead box Y (DBY) g
Human brain expres

PA (LUDW-) LUDWIG INST CANCER RES.

XX Martelange V, De Smet C, Boon-Fallieur T;

XX WPI: 1999-620430/53.

DR P-PSDB; AAY49636.

PT New nucleic acid encoding sarcoma-associated gene products, useful for
PT diagnosing, e.g. treating and preventing cancer -

XX Claim 1: Page 88-91; 93pp; English.

XX The present sequence encodes human sdpr.8 (HAG8), a sarcoma-associated
CC gene product (I). Agents, specifically sarcoma associated nucleic acids
CC (II) or their expression products that are tumour rejection antigens
CC (TRR), that selectively increase formation of HLA (human leucocyte
CC antigen)/(I) complexes are used for treating cancer, especially sarcoma
CC and carcinoma, in humans and other animals. Compositions containing
CC autologous cytolytic T cells (CTL), specific for the HLA/(I) complex,
CC are similarly useful, also transformed cells that stimulate such CTL
CC in vivo. (II) are also used: (I) as source of therapeutic antilense
CC sequences that reduce expression of (II); (II) for recombinant
CC production of (I); (III) particularly its fragments, as primers and
CC probes in usual hybridisation and amplification assays, for diagnosis,
CC prognosis and monitoring of tumours, or for measuring binding
CC specificity of HLA molecules or CTL clones; (IV) to identify related
CC sequences; and (V) for generating transgenic animals, e.g. for studying
CC cancer and immune responses to it. (I) are used to raise specific
CC antibodies (Ab) and therapeutically. Ab are used to diagnose tumours in
CC immunosays, also for delivering drugs, toxins, imaging agents etc. to
CC (I)-expressing cells.

XX Sequence 2365 BP; 775 A; 420 C; 592 G; 578 T; 0 other:

Query Match 100.0%; Score 2365; DB 20; Length 2365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ttggtaccgaagctcggatccctagtaacggccgcaagtgtctctggaagaagtggcgagga 60
DB 1 ttgtgaccgaagctcgatccctagtaacggccgcaagtgtctctggaagaagtggcgagga 60
QY 61 tagagagcgtggcggggggggcttagctctgfcggggtctcttaagagtgctgtggc 120
DB 61 tagagagcgtggcggggggggcttagctctgfcggggtctcttaagagtgctgtggc 120
QY 121 ttccctggcagcgtagctctacgacgtacaggtgggtgcagagctggaacggcaacg 180
DB 121 ttccctggcagcgtagctctacgacgtacaggtgggtgcagagctggaacggcaacg 180
QY 181 agctcgagcggccctctcttgaaacaatgtccacacaggaagaaatcccaagcctct 240
DB 181 agctcgagcggccctctcttgaaacaatgtccacacaggaagaaatcccaagcctct 240
QY 241 acgtggatcgttcgtagtcggcggaagctcgacagtgtcccgagcgcagagagagcgcg 300
DB 241 acgtggatcgttcgtagtcggcggaagctcgacagtgtcccgagcgcagagagagcgcg 300
QY 301 gcgagggagttgaaatcgaaacagctcctagagatatagtgtcgcgagaggtgtgtcgtcg 360
DB 301 gcgagggagttgaaatcgaaacagctcctagagatatagtgtcgcgagaggtgtgtcgtcg 360
QY 361 agagggacactctagcccccggagagcggtggcgctgtgtcaagagaaactgcgcgttgt 420
DB 361 agagggacactctagcccccggagagcggtggcgctgtgtgtcaagagaaactgcgcgttgt 420
QY 421 ttgtgcttgaagagcaacttgttggcgcgtaactcgctgtgtgtgtgtgtgtgtgtgtgtgt 480
DB 421 ttgtgcttgaagagcaacttgttggcgcgtaactcgctgtgtgtgtgtgtgtgtgtgtgtgt 480
QY 481 aatatacaagtaacaacaacccaatccaataataatacaagaacaacacagaatcat 540
DB 481 aatatacaagtaacaacaacccaatccaataataatacaagaacaacacagaatcat 540

DB 481 aatatacaagtaacaacaacccaatccaataataatacaagaacaacacagaatcat 540
QY 541 gtcaaaaatttttggcagcaaggcaatgcaaacgaaagcaaaagcagtgtatagaacattt 600
DB 541 gtcaaaaatttttggcagcaaggcaatgcaaacgaaagcaaaagcagtgtatagaacattt 600
QY 601 gtcaaaaagctgaagaagaatttaaatctagaatgtcggaattgtatctgtccaaact 660
DB 601 gtcaaaaagctgaagaagaatttaaatctagaatgtcggaattgtatctgtccaaact 660
QY 661 tctgtttggaagaatggaagcagacagataacaatgtgtgtgcagagatcggcaattgtata 720
DB 661 tctgtttggaagaatggaagcagacagataacaatgtgtgtgcagagatcggcaattgtata 720
QY 721 gatttggatcaaatatagagaaagggtttgaatgtgcaaaaacaaagtggcagatttata 780
DB 721 gatttggatcaaatatagagaaagggtttgaatgtgcaaaaacaaagtggcagatttata 780
QY 781 caaccataagaaaaaactttataaagagtcacatcgccacagatgtgccatgaagaata 840
DB 781 caaccataagaaaaaactttataaagagtcacatcgccacagatgtgccatgaagaata 840
QY 841 gaagcagatagcttggaagaaagaaatttaataacgttggatgtacttgaagatggg 900
DB 841 gaagcagatagcttggaagaaagaaatttaataacgttggatgtacttgaagatggg 900
QY 901 gaagaacgacatctcccaatccatctgacatgtgaacatttgaagcgtttcaatgttact 960
DB 901 gaagaacgacatctcccaatccatctgacatgtgaacatttgaagcgtttcaatgttact 960
QY 961 gaagttatggaanaaatatnaaagaagcaggttttcaaaaagccaacacatctagtcacag 1020
DB 961 gaagttatggaanaaatatnaaagaagcaggttttcaaaaagccaacacatctagtcacag 1020
QY 1021 gcatggcccatgtgtgttgcagaagatatagatcttatagaggttagcccgagatgaacagga 1080
DB 1021 gcatggcccatgtgtgttgcagaagatatagatcttatagaggttagcccgagatgaacagga 1080
QY 1081 aagaacatgtgttatatttaatgctcgtgattatcatctgtctctcaaccacgcttaaaa 1140
DB 1081 aagaacatgtgttatatttaatgctcgtgattatcatctgtctctcaaccacgcttaaaa 1140
QY 1141 ggtcagaagaatagaccggcgcatgttgatgtctaaccccaatcccggaattagacattcaa 1200
DB 1141 ggtcagaagaatagaccggcgcatgttgatgtctaaccccaatcccggaattagacattcaa 1200
QY 1201 gtgaagaggaatgtgtgcaaatatcaataaagggtctcggaagtgttgtgtatagtgt 1260
DB 1201 gtgaagaggaatgtgtgcaaatatcaataaagggtctcggaagtgttgtgtatagtgt 1260
QY 1261 ggtcgaatatagatgaaacaatatagaaagagcttaaaaaaggtgtgataatcatatgtgca 1320
DB 1261 ggtcgaatatagatgaaacaatatagaaagagcttaaaaaaggtgtgataatcatatgtgca 1320
QY 1321 actcccggaagaatgtaatgatctgcgcaaatgaaatgaactctgtcaatctgaaagaataaac 1380
DB 1321 actcccggaagaatgtaatgatctgcgcaaatgaaatgaactctgtcaatctgaaagaataaac 1380
QY 1381 tacttggctttatagatgaagcagacaagaatgttgaacatggaattgtgaacccagataatg 1440
DB 1381 tacttggctttatagatgaagcagacaagaatgttgaacatggaattgtgaacccagataatg 1440
QY 1441 aagaatttctgatagtgtgcgcccagataagtcagacaggtatagcaagtgctatactgtgct 1500
DB 1441 aagaatttctgatagtgtgcgcccagataagtcagacaggtatagcaagtgctatactgtgct 1500
QY 1501 catctagttcaatcgctctgcgcaaatcttatttgaagaacaaatgattgtctatgttgtgt 1560
DB 1501 catctagttcaatcgctctgcgcaaatcttatttgaagaacaaatgattgtctatgttgtgt 1560
QY 1561 acattgtatcattgttcgttaagtctcagtgaaagcaaaataataatgttaaccacggagaa 1620
DB 1561 acattgtatcattgttcgttaagtctcagtgaaagcaaaataataatgttaaccacggagaa 1620

QY	1621	gagaaatcggagctcaaatcgaatctttctacagatcgtatcgtatccacagacaagtcat	1680
Db	1621	gagaaatcggagctcaaatcgaatctttctacagatcgtatcgtatccacagacaagtcat	1680
QY	1681	gtcttcgcttcctcgaaaagcgtcttgcgagctccattcaatgaaatgaaatcgaat	1740
Db	1681	gtcttcgcttcctcgaaaagcgtcttgcgagctccattcaatgaaatgaaatcgaat	1740
QY	1741	atctcgtatgagctctctcgatcgtagatagaaagaagagatcggagaagcatagag	1800
Db	1741	atctcgtatgagctctctcgatcgtagatagaaagaagagatcggagaagcatagag	1800
QY	1801	aaactttaaaccgagcaaaagtgaagaataacaaatcgaatcgaatcgaatcgaat	1860
Db	1801	aaactttaaaccgagcaaaagtgaagaataacaaatcgaatcgaatcgaatcgaat	1860
QY	1861	gatcgtccatcgatcgtatcaacatcgtcttaatttgacttccacgaatatcgaagatc	1920
Db	1861	gatcgtccatcgatcgtatcaacatcgtcttaatttgacttccacgaatatcgaagatc	1920
QY	1921	gtcacacgaatcagcgccacagggaaagaagagagagactgtgttccatcgaacttg	1980
Db	1921	gtcacacgaatcagcgccacagggaaagaagagagagactgtgttccatcgaacttg	1980
QY	1981	actagaatcgtatcggaggtctgccttcgaatcgtatcaatcgtatcgtgaagaagcaatcag	2040
Db	1981	actagaatcgtatcggaggtctgccttcgaatcgtatcaatcgtatcgtgaagaagcaatcag	2040
QY	2041	agatctccagagagagcgtctgtatcaatcgtgtcgtgaaggtttgaagcacaatcaagcgaaag	2100
Db	2041	agatctccagagagagcgtctgtatcaatcgtgtcgtgaaggtttgaagcacaatcaagcgaaag	2100
QY	2101	gaaatcggaaagaataatcggaaagacctccaaggaagcccaagaattcatatgctct	2160
Db	2101	gaaatcggaaagaataatcggaaagacctccaaggaagcccaagaattcatatgctct	2160
QY	2161	cgtactacgtcgggtaagagatccaagaatttttgaagaatacgtgaagcagaatctgt	2220
Db	2161	cgtactacgtcgggtaagagatccaagaatttttgaagaatacgtgaagcagaatctgt	2220
QY	2221	gacatcgttcggcagctatcgaagagacgagcagatcttgacgactcttaaatatagtgct	2280
Db	2221	gacatcgttcggcagctatcgaagagacgagcagatcttgacgactcttaaatatagtgct	2280
QY	2281	tcgaataatagaaatcccgcttttcaactctcttaataaataagaatattttaaact	2340
Db	2281	tcgaataatagaaatcccgcttttcaactctcttaataaataagaatattttaaact	2340
QY	2341	aaaaaaaaaaaaaaaaaaaaaa 2365	
Db	2341	aaaaaaaaaaaaaaaaaaaaaa 2365	

XX	RESULT	2
XX	ABL13469	
XX	ID	ABL13469 standard; cDNA; 2210 BP.
XX	AC	
XX	AC	ABL13469;
XX	DT	
XX	26-MAR-2002	(first entry)
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 34889.	
XX		
KW	Drosophila; developmental biology; cell signalling; insecticide;	
KW	pharmaceutical; gene; ss.	
XX		
OS	Drosophila melanogaster.	
XX		
PN	WO200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		

PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-061A150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656660/75.
 XX P-PSDB; ABB69366.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PI genes from Drosophila and for elucidating cell signalling and cell-cell
 PR interactions -
 XX
 PS Claim 1; SEQ ID NO 34889; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB857737-AB872072), and the encoded proteins
 CC (AB857737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPAC
 CC at ftp.wipac.int/pub/published_pct_sequences.
 XX
 SO Sequence 2210 BP; 598 A; 509 G; 615 G; 488 T; 0 other:

[illegible]

RESULT 4
 AAST070200 ID AAST070200 standard; CDNA: 1098 BP.
 XX AAST070200;
 XX AAST070200;
 DT 13-FEB-2002 (first entry)
 DE
 CC DNA encoding novel human diagnostic protein #6004.
 OS
 KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 PN MO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001: 2001WO-US08631.
 PR 31-MAR-2000: 2000US-0540217.
 PR 23-AUG-2000: 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 DR P-PSDB: ABG06013.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1: SEQ ID NO 6004; 103bp. English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy technique
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and a
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS664197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1098 BP: 279 A: 306 C: 295 G: 218 T: 0 other:

Query Match	10.7%;	Score 253.2;	DB 23;	Length 1098;
Best Local Similarity	92.1%;	Pred. No. 6.6e-46;		
Matches 267;	Conservative	0;	Mismatches 23;	Indels 0;
				Gaps 0

```
QY      731 aaatagagagggaagtcttgaaatgccaataaaccaagtggcagaattaccaccaatta 790
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       5 aaatctaagaacgtgctgtgnaaacagaagaagaatbgttcgcattaccaccaatta 64
```

OY 791 agnaaaacttttaaaaggtccactgcacaaagtcacgttcaaaagtagaagcagata 850
 Db 65 agaaaacttttaaaaggtccactgcacaaagtcacgttcaaaagtagaagcagata 124
 OY 851 gtctgaggaagaanaatttaatacaacgtggagtaacttaagaatggaggagaacac 910
 Db 125 gtctgaggaagaanaatttttaataaacgtggagtaacttaagaatggaggagaacac 184
 OY 911 ctatcccaactctaccctgcacatttgatagcgccttcaatgttatctccgaggtcaagg 970
 Db 185 ctatcccaactctaccctgcacatttgatagcgccttcaatgttatctccgaggtcaagg 244
 OY 971 aaacacttaaaagagcaggttttcaaaagccaacacactatctcagtcacag 1020
 Db 245 aaacacttaaaagagcaggttttcaaaagccaacacactatctcagtcacag 294

XX	00	Sequence 4384 BP; 1235 A; 979 C; 1024 G; 1146 T; 0 other;
XX	01	at fdp.wipo.int/pub/published_pct_sequences.
CC	02	specification, but was obtained in electronic format directly from WIPO
CC	03	
CC	04	
CC	05	
CC	06	
CC	07	
CC	08	
CC	09	
CC	10	
CC	11	
CC	12	
CC	13	
CC	14	
CC	15	
CC	16	
CC	17	
CC	18	
CC	19	
CC	20	
CC	21	
CC	22	
CC	23	
CC	24	
CC	25	
CC	26	
CC	27	
CC	28	
CC	29	
CC	30	
CC	31	
CC	32	
CC	33	
CC	34	
CC	35	
CC	36	
CC	37	
CC	38	
CC	39	
CC	40	
CC	41	
CC	42	
CC	43	
CC	44	
CC	45	
CC	46	
CC	47	
CC	48	
CC	49	
CC	50	
CC	51	
CC	52	
CC	53	
CC	54	
CC	55	
CC	56	
CC	57	
CC	58	
CC	59	
CC	60	
CC	61	
CC	62	
CC	63	
CC	64	
CC	65	
CC	66	
CC	67	
CC	68	
CC	69	
CC	70	
CC	71	
CC	72	
CC	73	
CC	74	
CC	75	
CC	76	
CC	77	
CC	78	
CC	79	
CC	80	
CC	81	
CC	82	
CC	83	
CC	84	
CC	85	
CC	86	
CC	87	
CC	88	
CC	89	
CC	90	
CC	91	
CC	92	
CC	93	
CC	94	
CC	95	
CC	96	
CC	97	
CC	98	
CC	99	
CC	100	
CC	101	
CC	102	
CC	103	
CC	104	
CC	105	
CC	106	
CC	107	
CC	108	
CC	109	
CC	110	
CC	111	
CC	112	
CC	113	
CC	114	
CC	115	
CC	116	
CC	117	
CC	118	
CC	119	
CC	120	
CC	121	
CC	122	
CC	123	
CC	124	
CC	125	
CC	126	
CC	127	
CC	128	
CC	129	
CC	130	
CC	131	
CC	132	
CC	133	
CC	134	
CC	135	
CC	136	
CC	137	
CC	138	
CC	139	
CC	140	
CC	141	
CC	142	
CC	143	
CC	144	
CC	145	
CC	146	
CC	147	
CC	148	
CC	149	
CC	150	
CC	151	
CC	152	
CC	153	
CC	154	
CC	155	
CC	156	
CC	157	
CC	158	
CC	159	
CC	160	
CC	161	
CC	162	
CC	163	
CC	164	
CC	165	
CC	166	
CC	167	
CC	168	
CC	169	
CC	170	
CC	171	
CC	172	
CC	173	
CC	174	
CC	175	
CC	176	
CC	177	
CC	178	
CC		

Query Match	9.7%;	Score 229;	DB 23;	Length 4384;
Best Local Similarity	52.5%;	Pred. No. 1.9e-40;		
Matches 633;	Conservative	0;	Mismatches 495;	Indels 77;
				Gaps 3;

Dh	519	ttcaagctcagggctgagccagcttgccttcagtgagatggatatagttttggagtaagctcaaga	578
Oy	1070	ctggagaacgagaagacattggttatattttaaigtgcctggatattatattctgcgtccataac	112
Dh	579	ctggagctcgggaataacattatcttatttgcgtcgtccattgtacacataaacacacg	638
Oy	1130	ccagagcttaaaaggctcaaaaggaatgagccggcagctgtttgattctaacctccacttgggaat	118
Dh	639	caattcc---agagaaaggttgatgggcctatttgcgttggtgcgcgaacactcgaagac	695
Oy	1190	tggcacttcaagtaagaagaaatgcttgcacaatatt---catlaaaaggctctcgagctg	124
Dh	696	tggccacaacaggtggtgcgacaaagtgctctgaatatatgctgcagcttgcgttgaaactota	755
Oy	1247	tttggataatggttggtggaataatgagaatgaacaatatgaagagacttaaaaaaggtttag	130
Dh	756	cttgacataatggttgcgtcccaaaagaccacagatctcgatatttggaaaaggttg	815
Oy	1307	atatcataattgcactcccggaagattgaatgatctgcaaatgaaatgaaacttgcataac	136
Dh	816	aaacttgatgtgacaacactcgtgaagactgttgcactttttagagtttggaaaacacatc	875
Oy	1367	tgaagaatatatacctcttggttttataattgaacgacgaacaaatgcttggacatggatttg	142
Dh	876	tgaagagaacaaacttacccttgccttcttgatgaagcgtatagatgcttgatagagatttg	935
Oy	1427	aaccacgataatgaagaatttctgttaatgtgcgcgcgaataagcagacagtatagca	148
Dh	936	aaccacgaataaagaaaaatttcttgatccaataaagacctgtgaagcaacaacataatgtgga	995
Oy	1487	gtgtacatactgacctcatcttcatctgcgcgcgaacactctatcttgaagaacaacatga	154
Dh	996	gtgcacacttggccaanaagaaatgaagacagcttgcgtgaagatttccgaaagctatactt	105
Oy	1547	tgtctaatgttgtaatttgatcttagttgtgcgttaagtctcaaggagaagcaatatatttg	160
Dh	1056	atatcataattgtgcactgtgaacactgagtgacaacacaataacttccagatctggagtg	111
Oy	1607	taaacaccca---ggaagagaatctgagttcaaatgcaaacacttctacagagtatgtcat	166
Dh	1116	tatgtcatatgtctgcgaataaaggatgaaaagctattatctgtctatgagaagaatcatgttg	117
Oy	1664	ccagcg---acaaagtcacttctcttcttccgcgaanaagcgttgcgtgtaactatca	172
Dh	1176	agtaagaggaataaactattgttcttcttggaacaacaaagaagaatgtctgtaacttcca	123
Oy	1721	gttaccctaactcttggaataatatacatgtagagctctctgcabgagaatgaagaacaagag	178
Dh	1236	gaaaaatgaaagaaagatgtgggtgccttcccataggatcatccatgtgacaaggtcaagcag	129
Oy	1781	atcggagaaagacattatagagaaactttaaaccgcgaagaagtggagaatactaatgtcaactg	184
Dh	1296	aacgtgagctggttctaaatgaaatcaacaacatggaanaagcctcatcttgcattgtctacg	135
Oy	1841	atctagagctctagagagacttgatgtccatgaaogttacaacatgctctaatatttgacttc	190
Dh	1356	atgtgtgctcccaagaggtctatgagtgtgaaagatgtgaaatttgcatacaattatgataccc	141
Oy	1901	caacgaaatctagagaatctgataacacggaatggatgcgaacggaaagaagggggagctg	196
Dh	1416	ctaaactcccaagaggtattatcatctgatttggaagacttggaagactgcgtgaacaaacg	147
Oy	1961	gtgttccattaaactttgactgtaagaagatgtgaggtgtgcctctgattgtatata	202
Dh	1476	gacacgaatacacttctcttaccctataacaataaagacaagtgaagcacttactcttg	153
Oy	2021	ttcttgaagaagcaaatcagaatt	204
Dh	1536	tgtcttgcggaagctaatcaagaatt	1561

AA540962/c		
ID	AA540962	standard; cDNA: 2888 bp.
XX		
AC	AA540962;	
XX		
DT	17-DEC-2001	(first entry)
XX		
DE	cDNA encoding novel human enzyme polypeptide #178.	
XX		
KW	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;	
KW	liase; hyperproliferative disorder; immunodeficiency disorder;	
KW	autoimmune disorder; neurological disorder; metabolic disorder;	
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;	
KW	blood-related disorder; infectious disorder; gene therapy; cytostatic;	
KW	anti arthritic; nephrotoxic; anticoagulant; ss.	
OS	Homo sapiens.	
XX		
PN	WO20015301-A2.	
XX		
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01239.	
XX		
PR	31-JAN-2000;	2000US-0179065.
PR	04-FEB-2000;	2000US-0180628.
PR	24-FEB-2000;	2000US-0184664.
PR	02-MAR-2000;	2000US-0186350.
PR	16-MAR-2000;	2000US-0189874.
PR	17-MAR-2000;	2000US-0190076.
PR	18-APR-2000;	2000US-0198123.
PR	19-MAY-2000;	2000US-0205515.
PR	07-JUN-2000;	2000US-0209467.
PR	28-JUN-2000;	2000US-0214886.
PR	30-JUN-2000;	2000US-0215135.
PR	07-JUL-2000;	2000US-0216647.
PR	07-JUL-2000;	2000US-0216880.
PR	11-JUL-2000;	2000US-0217487.
PR	11-JUL-2000;	2000US-0217496.
PR	14-JUL-2000;	2000US-0218990.
PR	26-JUL-2000;	2000US-0220963.
PR	26-JUL-2000;	2000US-0220964.
PR	14-AUG-2000;	2000US-0224518.
PR	14-AUG-2000;	2000US-0224519.
PR	14-AUG-2000;	2000US-0225213.
PR	14-AUG-2000;	2000US-0225214.
PR	14-AUG-2000;	2000US-0225266.
PR	14-AUG-2000;	2000US-0225267.
PR	14-AUG-2000;	2000US-0225268.
PR	14-AUG-2000;	2000US-0225270.
PR	14-AUG-2000;	2000US-0225447.
PR	14-AUG-2000;	2000US-0225757.
PR	14-AUG-2000;	2000US-0225758.
PR	14-AUG-2000;	2000US-0225759.
PR	18-AUG-2000;	2000US-0226279.
PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226686.
PR	22-AUG-2000;	2000US-0227182.
PR	23-AUG-2000;	2000US-0227009.
PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	01-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	05-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	06-SEP-2000;	2000US-0230438.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.

Db 2278 TTAAGAGAGTAAAGATTTCACAGACTTGGAGCTGTGCTGCTTAGG 2219
QY 1259 gtggtggaataagatgaaacaataagaagacttaaaaaagtgatgatacaatg 1318
Db 2218 GAGGAACAGAAATCAGTACACAGATTGCTGCAATAAAGAGTGTGAATATTGTTT 2159
QY 1319 caactccggaagattgaatgactgcaatgagtaac-----tcgtcaatctga 1369
Db 2158 GCACACCTGCTGGAATGATGACATGTAGCCGCTAACAGTGTGGGTCACAAATCTTC 2099
QY 1370 agaataaactactggttttagaagcagacaagaatgttgaatgagttggaac 1429
Db 2098 GAAGAGTACATATGTTGTTTATGATGAAGACAGAAAGATTTTGACATGGTTTGAAC 2039
QY 1430 ccagataaagaagattgttagatgctccagataaggcagacagttatgaccagtg 1489
Db 2038 CCCAGTCATGCCATGCTGATTAATGTTGCTGATGACAGACGGTTATGTTTTCAG 1979
QY 1490 ctacatggccctcatcagttcatcgcctgcacaactctatttgaagaacaatgattg 1549
Db 1978 CTACTTTCCCAAGACTATGAGGCTTTGGCTCGCAGATCTCAGTAACCTATTGAAG 1919
QY 1550 tctatgtgttacatltggatcgtgctgtaagttcagtgtaagcaaatataatgttaa 1609
Db 1918 TACAAAGTTGGAGCAGAGAGTGTGTTGCTCAGAT---GTGAGACAAACAGATGTTGA 1862
QY 1610 ccacggaggagaagaatgagatcacaatgcaaaacttctcagaagatgctatccacag 1669
Db 1861 TTGAGAGAGAAAGAAATTTCTTGAACCTTCTGAGCTTCTAGGCCATTATCAAGAGCTCAG 1802
QY 1670 acaaaagatcgtctcgtctcgcgaagaagcgtgtcgagatcactatcaagtgaactaa 1729
Db 1801 GATCTGCTATTATTTTGTGGATTAACAGAACATGCTGATGCTCTTCTTAAGATTTGA 1742
QY 1730 tacttgaataatacatagtagtctctcgtcagagataagagaacagagatcgagaa 1789
Db 1741 TGAGAGCATCTTATCCCTTGGATGTCTTCATGAGGAGCATGATGATACAGAGATA 1682
QY 1790 aagcatagagaactttaaaacagcagaagtgagaatactaatgtcaactgatactagct 1849
Db 1681 GCATCAATAATGACTTTTAAGAAATGGGACCTGCACAACTTCTGTGGTACCTCTGTTGCTG 1622
QY 1850 ctagaagactgtagtccatgacgttacaacatgcttataatattggaacttccacggaata 1909
Db 1621 CCCGAGCTGATGATGAACATCATCTGATCTTGTATTAATTAAGCTGCCCAACCATT 1562
QY 1910 ttgaagaatacgtacaccggaatagggcgaaggaagaagacagggagagactggttcca 1969
Db 1561 ATGAGGATTATGTACACAGAGCGGCGGACTGGAAGAGCGAAGCAAGGCTTATGCTT 1502
QY 1970 ttacaacttgaactagaagaatgattgtaggggtgctcctcgaattgattaatatctgaa 2028
Db 1501 ATACTTTATCAGAGAGATCAAGCTCGCTATGCTGTGATCATATTAAGAAGCTCTTGA 1443

RESULT 8
AAS68838 standard: cDNA: 3247 BP.
ID AAS68838 standard: cDNA: 3247 BP.
XX
AC AAS68838:
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4642.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSQD INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
P-PSDB; ABG04651.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 4642; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 3247 BP; 1089 A; 551 C; 870 G; 737 T; 0 other;

Query Match 8.1%; Score 192; DB 23; Length 3247;
Best Local Similarity 51.3%; Pred. No. 2; 2e-32;
Matches 554; Conservative 0; Mismatches 510; Indels 16; Gaps

QY 962 aggtatggaataacatlaaaaggcaggttttcaaaagccacacactatcaagcaga 1021
Db 1289 agatctaaattccctccaagaagcatggtcatgaagaagccacccatccaagcaag 1348
QY 1022 catggccattgltgtgcaaggaatagatctatagagatagccacagctggaacagaa 1081
Db 1349 ctatccctgctataaagtctgtgacgagatgattggtcgtacaaacaggaagtga 1408
QY 1082 agaatctgttatatttaagtcctgattatcatctcgttccttaaccagccttaag 1141
Db 1409 agaacatlgctttctcgttgcctcatggttgaaca-----catatgatatcaaggtcattag 1465
QY 1142 gtcaagaataagaccggcagtgtagtcttaaccctccacccggaattagcaattcaag 1201
Db 1466 aggaaggaagaggccataagctgttcatcatgactccaactcgaaactgcttaacaga 1525
QY 1202 tagaagagaatgttgcaaatatcata---taaaaggtctcggaagtggttggtatag 1258
Db 1526 ttactaaagagtgtaagaagtcttccaagctttaggacttaagatggtctggttttaag 1585
QY 1259 gtggtggaataagatgaacaataagaagaagcttaaaaggtgtagatatcataatg 1318
Db 1586 gaggaacaggaatcagtgagcagatgtcgtagctgaagaaggtgctgaattatgttt 1645

Oy	1205	aagagaaatggttcgaataattccta---taagggcttcggagtggtttgtcatatgtg	1261
Db	1163	atccaaatgtaagcgggtcttggaaaagcatalaactcttcgatacgtgtgccgtatagag	1222
Oy	1262	gttgaaatgagatgaaacaaataagaagagcttcaaaaagagtgatgatalcataatggca	1321
Db	1223	gagggagatgtgvggagcaagcccaagcccttcaggaggggscgagatgtgtgtgta	1282
Oy	1322	ctccggaaagatgatagtatctgcgaatgagtaactcttcgaatcgtgaagaatatgaact	1381
Db	1283	ccccaggtccgactgtatagatcagatcagaagaagaagcaccacaactccaagagctctc	1342
Oy	1382	acttggcttttagatgaagcagacagaatgcttggacatgggatttgaaacccagataatga	1441
Db	1343	accttctgtttatbaagcagatctgaatcttcttgacatggatttgatacacaagtctgat	1402
Oy	1442	agatttctgttagatgtggccccagataagcagaacaggttatgacagatgtaactgtcc	1501
Db	1403	ccatagcaagatcatgttctgcttcgacacggcagacctcttatattagtgcaacttttgga	1462
Oy	1502	attcagttcatcgcctcgcaacatacttatcttgaagaacaaatgatatgtctatgtgtga	1561
Db	1463	agaagaatgtgaaaagtctggccccagagataccctctatcgaccctatctggagtgtgcagggag	1522
Oy	1562	catgtgatactatgtctgttaagtctcagtgaaagcaaaatataatgtataccaccggagag	1621
Db	1523	atatgtgagagggcaaatgaaatgtagacagagattgtgagatgtctccattcttcgaacta	1582
Oy	1622	agaaatgaggtctacatctgcaaaactttcttcaagaatgatagtcatctcacagaaaagctatg	1681
Db	1583	gtataatggaactctgcttaccgccgcgtctgttagaatctaacctcttcaggagtgctcc	1642
Oy	1682	tctctcttctcgaaaaagctgtgttgccgatacttctcaagtgcatactatcttggaaaata	1741
Db	1643	tctttcttctaataaaagccaaatctgcgaagagctagcgaataaacttaaacagaggggttc	1702
Oy	1742	tatccagttagatcgtctgtatgagaatagagaacagaagatccggagagaagaactagaga	1801
Db	1703	ataactcttggtcgtccacatgagcagatgtgatagtgacagaagctggaagaagaagatcttcag	1762
Oy	1802	actttaaacagcgcaagtgtgaataactaatctgcacactgatacgtcctctagaagactgt	1861
Db	1763	actttaagaanaaaggacatcccaagctctctgtgtgcacagaatgttgcgcccggtgtgtcgtg	1822
Oy	1862	atgtccatacgtctacaacatgtctataattgtacttccacgggaatattgaaagaactcgg	1921
Db	1823	aatctctctcaattaaagactgtctaatcaactatgatagtgtgcagggacattgatacgcaca	1882
Oy	1922	tacaacgaatatgggcgcacgcgggaagagcaggaagactcgtgttccatatacnaacttga	1981
Db	1883	ctcacagagatctggccgcacgaagaagacgggtgtgaaagaagtggtgtgcatatacctaacta	1942
Oy	1982	ctagaatgtatctgggggtgtccctctcctaaattgtatataattcttgaagaagcaaatcga	2041
Db	1943	ctcccaagacagcaaatcttctgtgtgacctgtctcggaacttggaaaggcccaatcaac	2002
Oy	2042	gtatcttcagagagactgtatcaatggc	2069
Db	2003	acgttctctaagaactcctatagatctgc	2030
RESULT 12			
AAZ88782			
ID AAZ88782 standard; cDNA; 771 BP.			
XX AAZ88782:			
XX			
XX 18-MAY-2000 (first entry)			
XX T. thermophila Hcl cDNA.			
XX Hcl: DEAD box protein; RNA helicase; cytosolic; cancer; diabetes;			

XX	OS	Tetrahymena thermophila.
XX	PN	WO200005388-A1.
XX	PD	03-FEB-2000.
XX	PF	10-JUL-1999; 99WO-EP04892.
XX	PR	22-JUL-1998; 98DE-1032783.
XX	PI	(AVER) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX	PI	Bonnet K, Huels C, Mueller S;
XX	DR	WPI: 2000-182709/16.
XX	DR	P-PSDB: AAY51577.
PT	PT	New Tetrahymena thermophila genes encoding RNA helicases, useful for
PT	PT	treatment of, e.g. cancer and autoimmune disease -
XX	PS	Claim 1: Page 42: 54pp; German.
XX	CC	This invention describes novel Tetrahymena thermophila genes encoding
CC	CC	RNA helicases (DEAD box proteins). The products of the invention have
CC	CC	cytostatic, immunoprotective and anti-diabetic activity and can be used
CC	CC	for gene therapy. The nucleic acids or RNA helicase polypeptides can be
CC	CC	formulated with a pharmaceutically acceptable agent and/or adjuvant for
CC	CC	treatment or diagnosis of cancer, autoimmune disease, in particular
CC	CC	multiple sclerosis or rheumatoid arthritis, Alzheimer's Disease, allergy,
CC	CC	in particular neurodermatitis, Type I or IV allergies, arthritis,
CC	CC	atherosclerosis, osteoporosis, acute and chronic infections and/or
CC	CC	diabetes and/or to influence cell metabolism, in particular
CC	CC	immunosuppression for all transplantation's and/or genetic diseases in
CC	CC	particular Werner Syndrome, Bloom Syndrome, Xeroderma pigmentosa and
CC	CC	connective tissue disorders. The nucleic acids can also be used to
CC	CC	identify functional interactive molecules or to find variants of RNA
CC	CC	helicases. The nucleic acids or proteins can be used to influence
CC	CC	protein biosynthesis, to inhibit or stimulate degradation of mRNA and/or
CC	CC	stabilize mRNA. The nucleic acids or polypeptides can also be used as
CC	CC	selection markers in molecular biology. The nucleic acid and polypeptide
CC	CC	can be heterologously expressed in useful plants. This sequence encodes
CC	CC	the Tetrahymena thermophila RNA helicase Hcl protein, described in the
CC	CC	method of the invention.
XX	XX	
SQ		Sequence 771 BP: 268 A; 113 C; 153 G; 237 T; 0 other;
QY		
		Query Match 7.7%; Score 182.2; DB 21; Length 771;
		Best Local Similarity 54.1%; Pred. No. 2.1e-30;
		Matches 417; Conservative 0; Mismatches 348; Indels 6; Gaps 2.
QY	1177	ccacatcggaattaggaactcaagtagaagaagaattggcaaat--atcatataaa 1233
DB	1	ccacatcagaacatcgctatcgtacaatcgaagaagagcgcgaagaattggttaactctc 60
QY	1234	gggctcggaagtgtctgtatattggtgtggaatatgagatgaacaataagaagcct 1293
DB	61	aagcttaaatgtgcttctatctatggtgtgctbacaataatccctaaagaagcactctc 120
QY	1294	aaaaaagtgtagatcacataatgtgaacatccgcgaagatctgatatctgcaatgagc 1353
DB	121	caataagtggtgtagtctgattgtactctcctgtagactattgactctttagaagc 180
QY	1354	aacttcgccaatcgaagaataaactactctgttttagatgaagaagcagaagatctg 1413
DB	181	gaactactactcttaagctagatgattactatctcgtattagatgaagcagatagaattta 240
QY	1414	gaacttgatttgaccaccgataatgaagaattttgtttagatctggtcccgataggcag 1473

Dh	241	gataatggcttcttggaaatttaattgaataaacttcgggttaattatgaaccgtatcgttaa	300
Oy	1474	acagttatgaaccagttgcbacatggtgcctacatcattacatgcctgcgcacaatcatt--	1531
Dh	301	acattatgtattcttcgcgtacccgtccctaagaatggttttaagaatcttggttaagaattatgc	360
Oy	1532	-tgaaagaaccacaaatgattgctctgtgttgtaaatggtacatggtctgaagtgtgtgaattccagt	1590
Dh	361	aagaaataccccggtttatgttccaatcgaagaaaccttgatattagctattaacgaagaatt	420
Oy	1591	aagcaaaataataattgttaaccaccgcggagagaatgagatggtacatgycanaactttcta	1650
Dh	421	aaataaattggtttatgtttacagatcatcataaagaanaaacaactacattacaagaatta	480
Oy	1651	caagatgtgtcatccacagacaagaatcattgctcttcgtcttcotcaanaagctttgcgat	1710
Dh	481	gattgttgcacttagaagaagatgattgcttgcgtcttaacaaagaagatgtgaa	540
Oy	1711	caattctcaatggaacctatactcttgtaaatataatcatgtagagctctgcgtgagatga	1770
Dh	541	agaaatgagtcgattatcttgaataaagaagaatttaagtgctctgttaccatggtgacaa	600
Oy	1771	gaacagagagatcgcggagaaagatctagaagaactttaaacaagccaagtgtgaaata	1830
Dh	601	gctctaaagaacagagactgttcttgaacaagatctcaaaagcggagatgacgaatcctt	660
Oy	1831	attgcacacgactgctgcctctagaagacttgatgtgccatgacgtttacaatgtctcata	1890
Dh	661	attgcctacagagatgacagaaagagtttggtgttlaagagtgctctccacgtatttaatt	720
Oy	1891	tttgactctccagaaatattgaagaatactgtaaccggaatagggcgacag	1941
Dh	721	taagattctccaaagttatggaagactctgcacaaatccgtttacaag	771

RESULT	13	
AAS54662		
ID	AAS54662	standard; DNA; 1464 BP.
AC	AAS54662;	
XX		
DT	13-FEB-2002	(first entry)
XX		
DE	Staphylococcus aureus DNA	for cellular proliferation protein #974.
XX		
AE	Antisense; ds:	prokaryotic cellular proliferation gene;
KW	antibiotic; antibacterial;	drug design.
XX		
OS	Staphylococcus aureus.	
XX		
PN	MO200170955-A2.	
PD		
XX		
PD	27-SEP-2001.	
XX		
XX		
PE	21-MAR-2001;	2001MO-US09180.
XX		
XX		
PR	21-MAR-2000;	2000US-191078P.
PR	23-MAY-2000;	2000US-206848P.
PR	26-MAY-2000;	2000US-207727P.
PR	23-OCT-2000;	2000US-242578P.
PR	27-NOV-2000;	2000US-253625P.
PR	22-DEC-2000;	2000US-257931P.
PR	16-FEB-2001;	2001US-269308P.
XX		
PA	(ELITR-) ELITRA PHARM INC.	
PI	Hasselbeck R, Ohlsen KL,	Zyskind JW, Wall D, Trawick JD, Carr GJ.
PI	Yamamoto RT, Xu HR;	
XX		
XX		
XX	WPI: 2001-611495/70.	
DR	P-PSDB: AAU36803.	
XX		

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
Claim 27; Seq ID No 8299; 51pp; English.
PS

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC [ftp.wipo.int/pub/published-pct-sequences](http://www.wipo.int/pub/published-pct-sequences).

Sequence 1464 BP; 540 A; 236 C; 300 G; 388 T; 0 other;

Query Match	7.4%	Score 174;	DB 23;	Length 1464;
Best Local Similarity	52.2%;	Pred. No. 1.6e-28;		
Matches 527; Conservative	0;	Mismatches 450;	Indels 33;	Gaps 5;

QY 988 ggtttcaaaagccaacacctatctagtcacagcatgcccattggttgcaagaata 1047
 || ||| || ||| ||||| || | | || |||||
 Db 4 ggaattaaagccgacacctatccaataaagacagtatccctatggtttacaagaatt 63

QY 1048 gatcttataggaqtaqcccaqactqgaacaqaaaacacattgttatttaatatgcctqga 1107

D_b 64 qatatccttqqcaaacqtaacaqtataaaacagaqaacattcgaatattccctta 123

1108 ttatctatctgaatccttcaaccacctaataaagatcaaaagaaatgaacccgacatatta 1167

Dh 124 attaaaaatattaaaaaaacaaaattcaatc-----gtr 162

Ov 1168 attctaactcccaactcagaaattatgcacttcaagtatgaagagaaatatttgcacaatatattca 1227

[illegible][illegible][illegible]

100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible][illegible]

1403 aaagatcgtcggacatcgggacatcgaacccagataaactgaagatcttcgtcttagatctgcgccca 1464

403 gaaatcgatcgaaatcagggatctcattcgatgatacgagatcttattatggaataaaatccagca 402

1465 g--atagcgagacagtatgaccagtgctacatgacctcattcagttcatcgccctcgca 1521

Db 463 gtacaacgtcaaacatgtgttctcagctacaatgcctaagaacatccaagctttagta 522

QY 1522 caatctattgaagaaccaatgtgtctatgtgtgtacatgtgatctagtctgta 1581

Db 523 caacaatttataaataccaataatcattaagacaatgataataatgaaatgtctgatcca 582

QY 1582 agttcagtgaaagcaaaataataattgttaaccacgcgaggaagaagaatgtagtccacatgcaa 1641

DB 583 caaatcgaaagatctacatcattgt---taagaattagagaatttcgatacttaca 639
OY 1642 acttctacagagatgcatccacagacaagtcattgtctctcttcctcgaagact 1701
DB 640 aattcctaga---tgctatcaacttgaattgacatcgattctcgagcgtacaacgt 696
OY 1702 gtcgagatcaacttcaactgaacttacttggaaatatatcagtagagtcctgcat 1761
DB 697 cgtgtgatgtaattacaagtgcttcttcttaagagataataaactgaagcttcat 756
OY 1762 ggaagatagagaacagagatcgaggagaacattagagaaactttaaacaagcagaatg 1821
DB 757 ggtgatattacacaagcagaacgctttagaagattataagaattttaaataatgaccacaa 816
OY 1822 agaatctaattgacaactgatactagcctctagagagacttgatgcatagagcttaca 1881
DB 817 aatatcttagcgtactatgattgacagcaagagactagataattctgtgtgagatcat 876
OY 1882 gctataatttgaacttccacggaatataagaagatacgtaaccggaatagagcgacag 1941
DB 877 gctataacttgaatactacctaagaatactgaaagctatacacacgctattgtctgacg 936
OY 1942 ggaagagcagagagagactggttccatctacacttgacacttgagaatga 1991
DB 937 gtcgtgtgtgtaagaagagtaacggtacggtacacttggtaaatcactga 986

RESULT 14
AAS51619
ID AAS51619 standard; DNA: 1530 BP.
XX
AC AAS51619;
XX
DT 13-FEB-2002 (first entry)
XX
DE Staphylococcus aureus DNA for cellular proliferation protein #36.
XX
KW Antisense; ds: prokaryotic cellular proliferation gene;
KW Antibiotic; antibacterial; drug design.
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001MO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX DR P-PSDB; AAU33760.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
XX PS Claim 27; Seq ID No 4201; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1530 BP; 571 A; 237 C; 309 G; 413 T; 0 other;

Query Match 7.4%; Score 174; DB 23; Length 1530;
Best Local Similarity 52.2%; Pred. No. 1.6e-28;
Matches 527; Conservative 0; Mismatches 450; Indels 33; Gaps 5;

OY 988 gttttcaaaagccacacactatcagtcacagagcattgtgttcaagaagata 1047
DB 121 gattttaaagacgcacacactatccaaaagacagatcccttagtcgttcaaggaatt 180
OY 1048 gatcttagagagtagccagactggaacagagaagacatggttatttaagtcctgga 1107
DB 181 gatatccttgaggcaagcctcaaacggtacaggaacagagagcattcgtgattcctta 240
OY 1108 ttattcatcgtgctcctcaaacccagccttaaaaggtcaaaaggaatagaccgcatgta 1167
DB 241 atcgagaagaagtagggagaacaaagggttcaatc-----gttg 279
OY 1168 gtcttaaccctccacgcgggaattagcaactcaagtgaaggaagatggtgcaaatatca 1227
DB 280 attttgacaccctacaagaagattgcaatgcaatgctgcaacaaattagaagaattagc 339
OY 1228 tata---aagggtcgcgaagtggtgtgtatgctgtgtggaatagatgaacaata 1284
DB 340 cgtgacaagaaggtgccaaagttgtactgttctggtgtgtatgcttcaagcacaat 399
OY 1285 gaagagcttaaaaaaggtgtgatcatcatatgcaccccggaagaattgaatgacgtg 1344
DB 400 aaggtccttgaaaaaaggcccaaatcgtgtgtggaacccctgggcgtgtatctgacacat 459
OY 1345 caattgagtaactcgtcaatcgtgaagataatacactctgttttagatgaagacac 1404
DB 460 ttaaatcgtgcacattaaaaacgagcaggaatcatacttgaatttagaagaagctgat 519
OY 1405 aagatgtgtgagatggaatttgaacccagataatgaagatttltgagtggtgcacca 1464
DB 520 gaaatgtgataatggtatctcatcgatgataatgatttattatgataaataatccagca 579
OY 1465 g---ataggcagacagtatgacacagtgctacatgctccatcagttcatcgctctgca 1521
DB 580 gtacaagcttgaagaacaaatgtgttcttcagctacaaatgcttcaagaatccaaagcttga 639
OY 1522 caatcttatttgaagaacaaatgtgtgtctatgtgtgtatctgtgatactgatactgtgctga 1581
DB 640 caacaatttatgaaatcaccaaaaaatacttaagaacaatgaataagaagaatgtcgaatca 699
OY 1582 agttcagtgaaagcaaaataatgtataccacagagaagaagaatgtgaatgcaatgcaa 1641
DB 700 caaatcgaaagaattctatacaattgt---taagaattagagaatcttgataacttaca 756
OY 1642 acttctacagagatgcatccacagacaagtcattgtctctcttcctcgaagact 1701
DB 757 aattcctaga---tgctatcaacttgaattgacatcgattctcgagcgtacaacacgt 813
OY 1702 gtcgagatcaacttacaagtgacttaacttggaaatatatcagtagagctctgcat 1761

Thu Jul 25 15:15:57 2002

us-09-923-831-42.rng

Page 17

OY	2009	aattgcttaatacttctgtgaagacaaatcaagatctccaggagagccttgatacaatgg	2068
Db	2306	aactgacctctgtgctcgagagagctcgccagaacgcgctgcamaagctctcttggatctcg	2365
OY	2069	ct 2070	
Db	2366	ct 2367	

Search completed: July 25, 2002, 12:25:05
Job time: 7485 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 09:11:00 ; Search time 4248.01 seconds

(without alignments)
11650.457 Million cell updates/sec

File: US-09-923-831-42

Perfect score: 2365

Sequence: 1 ttggtaccgagctcgatcc.....aaaaaaaaaaaaaaaa 2365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2365	100.0	2365	6	AR157514
2	2365	100.0	2365	6	AR173203
3	2365	100.0	2365	9	HS278110
4	2188.4	92.5	2200	6	AX086506
5	2188.4	92.5	2200	6	HS278110
6	934.2	39.5	179137	9	AC010282
7	934.2	39.5	170569	2	AC0108109
8	715.6	30.3	213447	2	AC083833
9	482.8	20.4	100061	1	AC016551
10	412.6	17.4	129261	2	AC068247
11	411	17.4	198052	9	AC019205
12	366.8	13.5	36864	9	AC010091
13	358	15.1	2370	3	AY051831
14	307.8	13.0	323	6	AR173200
15	307.8	13.0	1780	8	AB010259
16	267.8	11.3	1924	8	AF428313
17	266.4	11.3	2067	8	AY050375
18	266.4	11.3	2165	8	AY052351
19	266.4	11.3	2264	8	AY062745
20	261.6	11.1	1814	8	TOBRD810
21	257.2	10.9	2378	8	AF462866
22	256.2	10.8	2415	8	AY060589
23	235.2	9.9	2366	8	AC014595
24	229	9.7	65957	2	AC008094
25	229	9.7	170498	3	AC008093
26	229	9.7	181950	3	AE003677
27	229	9.7	227894	3	AE003677
28	228.2	9.6	135853	9	AC010971
29	227.6	9.6	2318	10	MMHEL1
30	227.6	9.6	2321	6	AX140497
31	226.4	9.6	202326	2	AL645849
32	224.6	9.5	2694	5	AF158370
33	222.2	9.4	2323	9	HSNP68M
34	222.2	9.4	2366	9	BC016027
35	222.2	9.4	2366	9	AY057700
36	221.8	9.4	2482	8	AC002337
37	220.8	9.3	82592	8	HSPE68A
38	220.6	9.3	2330	3	AF110009
39	217.6	9.2	2799	3	AF110007
40	217.6	9.2	5374	3	AF110008
41	212.2	9.0	5437	3	AF110008
42	211.4	8.9	2733	8	SCPE68
43	211.4	8.9	3740	8	SCYV112W
44	211.4	8.9	37800	8	SCCYV38K
45	211	8.9	3740	9	AB018344

ALIGNMENTS

RESULT	1	2365 bp	DNA	Linear	PAT 17-OCT-2001
AR157514	Sequence 42 from patent US 6245525.				
LOCUS	AR157514				
DEFINITION	Sequence 42 from patent US 6245525.				
ACCESSION	AR157514				
VERSION	AR157514.1	GI:16218459			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 2365)				
AUTHORS	Martelange/V., De Smet,C. and Boon-Fallieur,T.				
TITLE	Tumor associated nucleic acids and uses therefor				
JOURNAL	Patent: US 6245525-A 42 12-3UN-2001;				
FEATURES	Location/Qualifiers				
source	1..2365				
BASE COUNT	775 a 420 c 592 g 578 t				
ORIGIN	/organism="unknown"				

Query Match 100.0%; Score 2365; DB 6; Length 2365;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctggtaccagctcgagatccctagtaagcgccgctgctgctggaagaagctggcgga 60
DB 1 TTGGTACCGAGCTGGATCCCTAGTAACGGCCGCACTGTGCTGGAAAGAGGGCGGA 60
OY 61 taagaagcgtggcgaggggagctagcctcgtgcgggctccctlaagtaagcgctgagc 120
DB 61 TAGAGAGCTGGGGGGGGGGCTAGCTCTGGGGGCTCTTAAGTAGAGCGGCTGGTGGC 120
OY 121 ttccctgacagctagctcttaagcgtcaagctcaagttggtgaagagctggagcggaag 180
DB 121 TTCCCTGGACAGCTAGTCTTACGACGTACCGGTAGGTGTGACAGCTGGAGCGACAG 180
OY 181 acgtcggagcgccctctcttggaacaaatgcccacacgagagagctcccaagcctct 240
DB 181 ACGTGAGACCGCCCTCTCTTGGAACATGTCCACACGAGAGACCTCCCAAGCCCTCT 240
OY 241 acgtgggtcgtctgctagtcgggaagctcgagctgtcccggaagcgccgaagagggcg 300
DB 241 ACGTGGGTCTGCTGTCTAGTGGCGGAGCTCGACAGTGTCCGAGCGCCAGAGAGAGCG 300
OY 301 gcggaggaagtcgaatcgaaacaggtcccgaggaatagtgctcgcaaggtggtcgctg 360
DB 301 GCGGAGGAGTGAATGGAACAGAGTCTCGAGGATATAGTGTGCGAGAGTGTGCGCTGG 360
OY 361 agagagacacctagagcccgaggagcgctggcgctgtgtcaagaggaagctccctgtgt 420
DB 361 AGAGGACCTCTAGGCCCCCGAGGCGCTGGCCCTGTGTACAGAGGAACTGCCGTGTGT 420
OY 421 ttgctcttgaaagacaccttgcttgccggttaactgctgctgctgctgctgctgctgct 480
DB 421 TTGCTCTTGAAAGACCACTTGTGTGCGCGGTATCGTGTGCTGCTGCTGCTGCTGCTGCT 480
OY 481 aatataaagctacaacaaac 540
DB 481 AATATACAAAGTACAAACAAACCAACCAATTCCAATTAATACAGAACACCAACCAATCATTA 540
OY 541 gtcaaaactcttgccagcaagcaatgcaaaacgaagcaaaacagtgatgaacaattc 600
DB 541 GTCAAAATTTTGTGGCAAGCAAGCAATGCAAAACGAAGCAAAACCACTGTATGACAAATTTT 600
OY 601 gctaaaagctagaagaac 660
DB 601 GTTAAAGCTAGAAAGAAATTTACAAATTCAGAAATGCGAATGTATCTGCANTCCAACT 660
OY 661 tctgttggaagaagctggaagcagatacaaatgctgttgcaagagatcgccatgata 720
DB 661 TCTGTGTGAAAGATGGAAGCACAGATTAACAATGTTGTGTCAGAGATCGGCCANTGTATA 720
OY 721 gattggagatcaaatagagaaggaaggttgaaatgcaaaacaaagctggcgatgata 780
DB 721 GATTGGATCAATATAGAGAGGAGGTTGAAATGGCAAAAACAAAGGGGCGAGATTTTA 780
OY 781 cccacaaatgaagaac 840
DB 781 CACCAATTAAGAAAACTTTTAAAGAGTCCACCTGCCCAAGTGTCCATGTGCAAAAGTA 840
OY 841 gaagcagatagctggaaggaagaacacacacacacacacacacacacacacacacac 900
DB 841 GAAGCAGATAGTGTGGAGAAAGAAATTTTAATTAACGTGGGATGACTTGAAGATGGG 900
OY 901 gagaagac 960
DB 901 GAGAAACGACCTATCCCAATCTTACCTGCACATTTGATGAGCCCTTTCAATGTTATCTT 960
OY 961 gaagctatggaagaac 1020
DB 961 GAGGTTATGGAAGAACTTAAAGAGCAGGTTTCAAAAGCAACACACTTTCAGTCAACAG 1020
OY 1021 gcatggccacatgctgtctgcaaggaatagatcctatagagtagccagactggaacaga 1080
DB 1021 GCATGGCCATGTGTTGCAAGGAATAGATCTTATAGGAGTAGCCAGACTGGAAACGGA 1080

OY 1081 aagaacatgctgctatgaatgctggaattatcaatcgtgctcccaaccacacacacac 1140
DB 1081 AAGACTGTGTGTTATTAAATGCTGGAGTTATTCATGTGCTTCAACCAACCTTAA 1140
OY 1141 gctcaaggaatagaccggcagatgctagctcctaactcccaactcggaaatagacactcaa 1200
DB 1141 GCTCAAGGAATAGACCGGCATGTTAGTCTTAACCTCCACCTCGGAATTAGCACTTCA 1200
OY 1201 gtgaagagagaatctgcaaatctcataaaggctcgggaatgctgttgaatgata 1260
DB 1201 GTGAAGAGAAATGTTGCAAAATTTATTAAGGCTTCGAGATGTTTGTGATATGGT 1260
OY 1261 gctggaatagagatgaaacaaatagaagcctlaaaaaagctgagatcatcaatctgca 1320
DB 1261 GGTGGAATAGAGATGAACAAATAGAAAGCTTAAAGAGGTGTATGATATCATTAATGCA 1320
OY 1321 actccggagaagatgaaatgcatcgcgaatgagtaactcgtcaactcgaagataaac 1380
DB 1321 ACTCCGGAAGATTCATGATGATCTGCAATGAGTAACTTCGTCAATCGAAGATATMACC 1380
OY 1381 tacttgctttagatgaaagcaagaatgctgacatgggaattgaaacccagataatg 1440
DB 1381 TACTTGCTTTAGATGAAGCAAGCAAGATGTTGACATGGGATTTGAACCCAGATATG 1440
OY 1441 aagatttgcttagatgctgcccagataagcagacagctatgacagtgctatacgct 1500
DB 1441 AAGATTTGTTTATGATGTGCCCCAGATAGCGACAGATTAATGACCACTGCTAATGGCT 1500
OY 1501 catcaagctacgcctcgac 1560
DB 1501 CATTCAGTTCATCCCTCGCACAACTTATTTGAAAGAAACCAATGATTTGCTATGTGGT 1560
OY 1561 acattgagatcagctgctgctgaagctcagtggaagcaaaatataatgtgaacacagga 1620
DB 1561 ACTTTGATCTTAGTCTGCTGTAAGTCAAGTGAAGCAAAATATATTTGAACACCGAGAA 1620
OY 1621 gagaatggagatcacatgcaaacctctctacagatgctcatccacagaacacatcat 1680
DB 1621 GAGAAATGAGATCAATGCAAACTTTTTCAGAGATATGATCATCACAGCAAAAGCATTT 1680
OY 1681 gctctgcttctcgaagaagcgtgtgcaagctacacacacacacacacacacacacac 1740
DB 1681 GCTTCTGTTTCGAAAGCGTTGGCGGATCATTAACATGACTATTAATCTTGAAT 1740
OY 1741 atatacgtagaagctctgcacatggaatagaagaacaggaagctggggaagaacataag 1800
DB 1741 ATATCAGTAGAGTCTCTGCATGAGATAGAGAACAGAGATCGGAGAAAGCATTAAG 1800
OY 1801 aactttaaagaagcaagctggaatacctaacttgcaactgcatctagcctlaagaagct 1860
DB 1801 AACTTTAAACAGCAAAAGTGAAGATATTAATTCACACTGATTAAGCTTACAGACTT 1860
OY 1861 gattgcacatgacgttaacacatgctataattgagcttccaggaatataagaagatc 1920
DB 1861 GATGTCCATGACGTACACATGTCTAATATTTGACTTCCAGGAAATATGAAGATTC 1920
OY 1921 gtaacccaatagagcgcaaggaagaagcaaggaaggaagctgtgttccatlaaacactg 1980
DB 1921 GTAACCCGAATAGGCGCACAGGAGAGAGGAGAGACTGTGTTCATTCAATCAACTTG 1980
OY 1981 actagaagaatgtaggaaggtgtgcctctgaaatgataatctcggaaagagaacacag 2040
DB 1981 ACTGAAGATGATTTGAGAGGTTGCTTGAAATGATTAATTTGGAAGCAAAATTCAG 2040
OY 2041 agtattccagaagagctgtgatalcaatgctggaaggtttggaacacacacacacacac 2100
DB 2041 AGTATTCAGAGAGGCTTTGATCAATGGCTGAGAGGTTGAGGCACTCAACGGAAGAG 2100
OY 2101 gaaatggaaagaaataatggaagaacacacacacacacacacacacacacacacacac 2160
DB 2101 GAAATGGAAAGAAAAATGGAAGACCTCAAGAGAGGCCCAAGAGATTTCATTAAATGCTTT 2160

```
QY 2161 ctgtactagtgaggtagaagatttcttagaataatgaagacagaagatg 2220
|||||
Db 2161 CTGTACTAGTGGGGTAGAGATTCTTTTGAATATAGTAAGACAGATATG 2220
QY 2221 gacatgttgacagatgaagaacgagactgatttgactgattcttaataatagtg 2280
|||||
Db 2221 GACATGTTGGCAGATGATAGACACCGACTGATTTGACTGATCTTAATATATAGTGT 2280
QY 2281 tgaataataagaatccagtgtttatacttcttaataataagaatatttaact 2340
|||||
Db 2281 TGAATAATATAGATCCAGTGTATTATCTTTTAATATAAATAGATATTAACTT 2340
QY 2341 aaaaaaaaaaaaaaaaaaaaaa 2365
|||||
Db 2341 AAAAAAAAAAAAAAAAAAAAAA 2365

RESULT 2
LOCUS ARL173203 2365 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 42 from patent US 6303756.
ACCESSION ARL173203
VERSION ARL173203.1 GI:17912694
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2365)
AUTHORS Martelange Y., De Smet C. and Boon-Fallieur T.
TITILE Tumor associated nucleic acids and uses therefor
JOURNAL Patent: US 6303756-A 42 16-OCT-2001;
FEATURES
source 1. 2365
BASE COUNT 775 a 420 c 592 g 578 t
ORIGIN

Query Match 100.0%; Score 2365; DB 6; Length 2365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgtaccagctcgatccctagtaacgacgcagctgtctgaaagatggcgga 60
|||||
Db 1 TTGTAACGAGCTCGAATCCCTAGTAACGCGCCAGTGTGCTGGAAGATGGCGGA 60
QY 61 tagaagcgtaggggggggtagcctcgtagcggtcccttaagtagcggtcggtggc 120
|||||
Db 61 TAGAGAGCTGGGGCGGGGGGCTTAGCTCGTGGGGCTCTTAAGTAGCGGCTGCCGTGC 120
QY 121 ttccttgcaacgtacttcaacgacgtcaacgtcaggtgtagtgcagagcttgaaacgcaacg 180
|||||
Db 121 TTCCTGCGACGCTAGTCTTACGACGTCAACGCTCAGGTGAGTGTGSCAGAGCTGAGCGGCAAG 180
QY 181 acgtcgagcgccctctcttgaaacaatgtcccaacgagagatcccaagcctct 240
|||||
Db 181 ACCTCGAGCGCCCTCTCTTGAAACAATGTCCCAACAGGAGAGCTCCCAAGGCTCTCT 240
QY 241 acgtggtcgttctagtcagcgaagctgcacagtgctcccgagcgccagagagagcgcg 300
|||||
Db 241 ACGTGGTGTCTTCTAGTCTGCGGAGAGCTGACAGTGTCTCCGAGCGCCAGAGAGAGCGCG 300
QY 301 gcgagagagatgaatcgaaacaggtccctgaggaatatagtgttcgcgagagtggtcgctgg 360
|||||
Db 301 GCGAGAGATTTGAATCGAAGAGTCTGAGAGGATATAGTGTGCGAGAGGTGCTGCTGG 360
QY 361 agagacacctctagcccccgagagcgctgacgctgctgcacagagaaactgcgctgtg 420
|||||
Db 361 AGAGGACACTTAGGCCCCCGAGAGCGCTGGCGCTGCTGACAGAGAACTGCGCGTGTGT 420
QY 421 ttgtgcttgaagagcgaacttgttgcgcggtgaatcgtgctggtgggtcaaaaaataaag 480
|||||
Db 421 TTTGCTTTTGAAGGCACTTGTGTGGCGCGTATGCTGCTGCTGCTCAAAAAATTAAG 480

QY 481 aatatacaaggtacaacaacacacacatccaaatatacaagaacacacacgaatcata 540
|||||
Db 481 AATATACAAGGTACAACAACACACACATCCAAATATATACAGAACACACAGAAATCATTTA 540
QY 541 gtcaaaaatttttggcagcaaggcactgcaaaacgaagcaaaagcagtgatgacaaatttc 600
|||||
Db 541 GTCAAAATTTTTTGGCAGCAAGGCAATGTCAAAACGAAACCAAAACAGCTGATGACAAATTTT 600
QY 601 gttaaaagctagaagaataatatacaatcagaatgcggaattgatactgcatlccaaact 660
|||||
Db 601 GTTAAAGCTTAGAAGAAAATTAACAATTCAGAAATGCCGAATTCATCTCATTTCCAACT 660
QY 661 tctgttgaaagaatggaagcgacagatacaatgtgtgtgcagagagatcgccatctgata 720
|||||
Db 661 TCTGTGTGAAAGATGGAAGCGACATATATCAATGTGTGTGTCAGAGATCGCGCATTTGATA 720
QY 721 gatcggatcacaatagaagaaggtctgaaatgtagcgaataaaacaaagtggcagattcta 780
|||||
Db 721 GATTGGATCAAAATTTAGAGAGAGGATTTGAAATGGCAAAACAAAGTGGGCAATTTTA 780
QY 781 ccaccaatttaagaataactttataaagaagccacgtgcgcaagtgccatgtcacaagta 840
|||||
Db 781 CCACCAATTAAGAAAACCTTTTATTAAGATGCCCTGCCAAGTCCCATGTCAAAAGTA 840
QY 841 gaagcagatagttgagagaaagaataatcttaataaacytggatgacttgaagatggg 900
|||||
Db 841 GAACGACATAGTTTGAAGAAAGAAATTTTAATTAACGTTGATGATCTTGAAGATGGG 900
QY 901 gagaagaacacctcccccaatccctacgtccacatttgatgaagcctttaaagtatcct 960
|||||
Db 901 GAGAAAGACCTTATCCCAATCTTACCTGCACATTTTATATACCCCTTCAAAATTTATCT 960
QY 961 gaagttatggaataacataaaaggcagggttctcaaaagcacaacctatcagtcacag 1020
|||||
Db 961 GAGGTTATGGAACATTTAAAAAGCAGAGTTTCAAAAGCAACACCTTATCAGTCACAG 1020
QY 1021 gcatggcccatgtgttgcagaagaaataatcttaagaaatagccaagcttgaaacagga 1080
|||||
Db 1021 GCATGGCCCATGTGTGTGCAGAGAAATATATCTTTAGAGATATACCCAGATGAAACAGGA 1080
QY 1081 aagacattgttactttaaactgactgactttaaacttgccttcaaacccagaccttaa 1140
|||||
Db 1081 AAGACATTTGTTATTTAATGCTCGATTTATTCATCTGCTTCCTCAACCCAGCTTAAA 1140
QY 1141 ggtcaaaagaaatgaccggcgatgttagtttaactcccaactggaaatgaactttaa 1200
|||||
Db 1141 GGTCAAAAGAAATGACCCGGCATGTTAGTCTTAATCTCCACTCGGAAATTTACACTTCAA 1200
QY 1201 gtagaagagaaatgttgcacaataatcataaagaaggtctcgagatgttctgttataatgt 1260
|||||
Db 1201 GTAGAAGGAGAAATGTTGCAAAATATTCATATAAGGCGCTGCGAGTGTGTATATAGT 1260
QY 1261 ggtgaaatagagaatgaataaataatagaagagctttaaagaaggttaaatatcataatga 1320
|||||
Db 1261 GGTGAAATAGAGATGACAAATAGAAAGAGCTTAAAAAGGTGAAATATCAATTAATGCA 1320
QY 1321 actccggaagaaatgaaatgactgcacaatlgagtaacttcgccaacttgaagaataaacc 1380
|||||
Db 1321 ACTCCGGAAATGAAATGATCTGCAAAATGATGAACTTCTCAATCTGAAAGAAATTAACC 1380
QY 1381 tacttggttttaagaatgaagaagaagaaggtgttgaaacttgagatttgaacccaagata 1440
|||||
Db 1381 TACTTGTTTAAAGTGAAGAGCAAGATGTTGACATGGGATTTGAACCCAGATTAATG 1440
QY 1441 aagaatttgtatagatgtgcgcccaataggaagaacagttatgacaaagtgtcataatgacct 1500
|||||
Db 1441 AAGATTTTGTAGATGTGCCCCAGATAGGCAAGACAGTATTAAGCCAGTCTCATAGGCTT 1500
QY 1501 catcagttatgcgcctcgacaacatcttatattgaagaagaacaaatgatttctaatgttgt 1560
|||||
Db 1501 CATTCAGTTATGCCCTCGACACATCTTATTTGAAGAGCAACATGATGTCTATGTGTGT 1560
```

QY 1561 acattgga tctagttgctgtaagttcagtgagcaaaatataatgtgaaccacagga 1620
|||||
Db 1561 ACATTGATCTAGTGTCTGTAATTGAGGAAGCAAAATATATTTGACCCAGGAA 1620
QY 1621 gggaaa tggagtcacatgcacatcttctacagagtalgtcatcacagaagtcac 1680
|||||
Db 1621 GGAATATGATCATCATGCAAACTTTCTACAGATATGCTATCCACAGCAAGTCATT 1680
QY 1681 gctctgcttctcgaaaagctgttcgga tcaactatcaagtgacctaacttggaat 1740
|||||
Db 1681 GCTCTGCTTCTCGAAAAGCTGTGCGGATCCTTATCATAGTACCTATACTTGAAT 1740
QY 1741 atatcagtaaggtctctgcatacagagagagagagagagagagagagagagag 1800
|||||
Db 1741 ATATCATGATGAGTCTCTGCAATGAGATAGACAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 aacttcaaaacagcaagtgagataactatgcacatgcacatgcacatgcacatgc 1860
|||||
Db 1801 AACTTTAAACAGGCMAAGTGAGAACTGATGCACTGATCTAGAGACTT 1860
QY 1861 gatgtccatgcagttacacatctctataatcttgcacatgcacagagatgcagaat 1920
|||||
Db 1861 GATGTCCATGACGTTACACATGCTATTAATTTTACCTTCCACGGAATATGAGAAATAC 1920
QY 1921 gtacacagtaagagcgacagagagagagagagagagagagagagagagagagag 1980
|||||
Db 1921 GTACACCAATAGCGCGCACGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 actgagaatgtatggagaggttgcctcgaatgtatgtatgtatgtatgtatgtat 2040
|||||
Db 1981 ACTGAAATGATGAGAGGCTGCTCTGTAATGATTAATTTCTGGAAGACCAATACAG 2040
QY 2041 agtattccagagagagctgtatcaatctgagagagagagagagagagagagagag 2100
|||||
Db 2041 AGTATCCAGAGAGCTGTGATCAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 gaaatggaagaaagaaatgtgaagaaacctcaagaaagcgcaagaaagttcaata 2160
|||||
Db 2101 GAAATGGAAGAAAGAAATGGAAAGACCTCAAGAGAGCGCCAGAAAGTTTCATTAATGCTTT 2160
QY 2161 ctgtactagtgaggtagagaaatccaagattcttgaataatagtaagaacagaat 2220
|||||
Db 2161 CTGTACTAGTGGGTAGAGAAATTTTAAAGATTTTAAAGATTTTAAAGACAGACTAT 2220
QY 2221 gacatgttcgacagtagaagagagagagagagagagagagagagagagagagag 2280
|||||
Db 2221 GACATGTTGGCAGTATGAAGACCGGACGATTTGACTGATTTCTTAAATATATGCTT 2280
QY 2281 tgaataatagaaatcagatgttttaacttctttaaataaataagaatlatlaaact 2340
|||||
Db 2281 TGAATAATAGAAATCCAGTGTTTTATACTTCTTAAATAAATAGAACTTAAACTT 2340
QY 2341 aaaaaaaaaaaaaaaaaaaaaa 2365
|||||
Db 2341 AAAAAAAAAAAAAAAAAAAAAA 2365
RESULT 3
HSA278110 2365 bp mRNA linear PRI 26-SEP-2000
LOCUS HSA278110
DEFINITION Homo sapiens mRNA for DEAD-box protein (HAGE gene).
ACCESSION AJ278110
VERSION AJ278110.1 GI:8216984
KEYWORDS dead-box protein, HAGE gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2365)
AUTHORS Martelange V., De Smet C., De Plaen E., Lurquin C. and Boon T.
TITLE Identification on a human sarcoma of two new genes with
tumor-specific expression
JOURNAL Cancer Res. 60 (14), 3848-3855 (2000)

★

MEDLINE 20374312
REFERENCE 2 (bases 1 to 2365)
AUTHORS Martelange V.M.F.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Martelange V.M.F., Ludwig Institute for
Cancer Research, Universite Catholique de Louvain, Av Hippocrate,
74, 1200 Bruxelles, BELGIUM
FEATURES
source
1. 2365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q12-13"
/cell_line="LB23-SAR"
/cell_type="rhabdomyosarcoma"
/note="marker GATTA1808-D6S284"
208. 2154
/gene="HAGE"
208. 2154
/gene="HAGE"
208. 2154
/function="putative RNA helicase ATP dependent"
/product="DEAD-box protein"
/protein_id="CAB92442.1"
/db_xref="GI:8216985"
/translation="MSHGGAGKASTVWVVASRSTYSRAPERRAEELNTRGEGTS
VGRGRRRTSRPEPVAAAGHELDPLCFALSKSEVAVIRGSKIKNTOSTNTTID
I10DPESLVKIFGSKAMOTKRAVIDNFVAKLEENNSBCGIDTAPFPKDGSTTD
NNVAVGDRPLIDMDI1REGLKAMOTKRAVIDNFVAKLEENNSBCGIDTAPFPKDGSTTD
ENFNI1WDDLDKDEKREP INPPTCFDDAFQCEPVEMENIKKAGFPPIQSOAMPY
LOGIDLIGVATGOTGKTLCLYLPMPG IHLVLPISLQGRNPGMLVLTPLRELALQVGS
ECKSYKGLRSVYGGGGRNDEQIEELKGVDDI1ITPRLNDLQMSNPNLNTY
LVLEADKMLDMGFEPOIMK ILLDVRPROMVMSATPMSYHRLAQSLKEPMIVY
GLDILVSSVSKRONITITTEERKSMHMTPELOSMSSTDYIVYVRKAYADLSDDI
LGNISVSLGDEQREDRKALENKTKKRNILATDLSRGLDHDVHTYNDPFR
NIEEYVIRIRTRAGRTVSITTLTRNDMNVASSELINILIRANOSIRPELVSAER
EAHORKEKRMERPOGPKKEH
BASE COUNT 775 a 420 c 592 g 578 t
ORIGIN
Query Match 100.0%; Score 2365; DB 9; Length 2365;
Best Local Similarity 100.0%; Pred No. 0;
Matches 2365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgggtacagagctcgatccctagtaagcgccagtgctgtgaaagagtgggcgga 60
|||||
Db 1 TTGTTACCGAGCTCGATCCCTACTAAGCGCCGCACTGCTGGAAGAGTGGCGGGA 60
QY 61 tagagagcg tgggaggggagctagcctctgtcgagcctcccttaagtagcgctcg tggc 120
|||||
Db 61 TAGAGAGCGTGGCGGGGCGGCTAGCCTGCTGCGGGCTCTTAAGTAGCGGCTGCGTGC 120
QY 121 ttccttgacagctagttcttaagcagtcagtgagtgagtgagcagtcagtcagcag 180
|||||
Db 121 TTCCTTGACAGCTAGTCTTACGACGTCACGTCAGTGTGCTGACAGACTGAGCGCAAG 180
QY 181 acgtcgagcgccccctcttggaacaatgtccaccacagagagagctcccaagagctct 240
|||||
Db 181 ACGTGCGAGCGCCCCCTTTTGGAACAATGTCCACACGAGAGAGCTCCCAAGGCTCT 240
QY 241 acgtggagctgtgtagtcagcgagagctcgacagtgctccgagcgccagagagagcg 300
|||||
Db 241 ACGTGGGCTGTGTTGCTAGTGGCGAGAGCTGACAGTGTCCGACGCGCAGAGAGAGCGCG 300
QY 301 gcgagaggttgaaatcgaaaggttccctgagggagatagtgtcggcagaggtgg tgcctg 360
|||||
Db 301 GCGGAGAGATTGAATCGAACAGGCTCTGAGAGATTAATGTCGCGACAGAGTGGTCCCTGG 360
QY 361 agaggaaccttaagcccccgagagcgctgtgcagcttggtcagaggaacttcgcctgtgt 420
|||||
Db 361 AGAGGACCTCTAAGGCCCCCGAGGCGCTGCTGTCACAGAGAAATGCGCGCTGTGT 420

[illegible]

Dp	1501	CATTGAGTTTCATCGCCTGCACATCTTATTGTGAAGAACAATGATTGCTTATGTTGGT	1560
Oy	1561	acatttgatctagtctgtcgtgaagttcaatgtaagaacaaataataatgtaaccacccggaggaa	1620
Dp	1561	ACATTGTGATCTATTGGCTGTAACTTCAGTAAGCAAAAATVTAATVTGTAAACCAGCAAGAA	1620
Oy	1621	gaagaaa ttgagtcacaatgcacaaacttctcacagatctg tcatcacacagacaagaatcact	1680
Dp	1621	GAGAAATGGAGTCACATATGCACAACTTTTTCACAGAGATGTATGTCATCCACAGACAAAGTCATT	1680
Oy	1661	gtcttcgtttcttcgaaagaactgtctcgagatcactatccaagtgacctaatatcttggaaat	1740
Dp	1661	GTCCTGGTTTTCTCGAAAAAGCTGTTGGGATCCATTATCAAGTGACCTTAATVACTTGGAAAT	1740
Oy	1741	atalcagtagaaccttcctgcgca ttgagaa taagaaacagagaatctgggagaaagca cttaga	1800
Dp	1741	ATAATCATGATAGTAGTCTCTGCAATGAGATAGATAGAACAGAGAGATGGGAGAAAGCAATTAGAG	1800
Oy	1801	aacctttaaacacgggaaagygaaataactaatctgcaactatctagcctctcaggagactt	1860
Dp	1801	AACCTTTAACAACGSCAAAGTGAGAAATACATTAATTTGCAACTATCTATACCTCTTAGAGACTT	1860
Oy	1861	gagtcacatgaaacttatcacatctgctcta taaatttgaacttctccagagatatgtgaagaa tac	1920
Dp	1861	GATGTCCATGACGATTCACATAGCTCTTAATATTTTGTACTTCCACAGAAATATTTGAAGAAATAC	1920
Oy	1921	gtccaccgaataaggagcgacaggaaagaagagagagagactgtgtttccattacaaacttg	1980
Dp	1921	GTAACCGAATAGGGCGACGCGAABAGAGAGAGAGAGACTGTGTCTTTCCATTNNCAAATCTTTG	1980
Oy	1981	actagaatactgatctggagggctgtgcctcttgaaattgataataatcttgaaaaagcaaatcag	2040
Dp	1981	ACTGTAAATGATTTGGAGGGTGTGCCCTCTGAAATGATTAATTTCTGGAAAAAGCAAAATCAC	2040
Oy	2041	agatattccagaaagaaactgtgtataactatgcttgagagagtttgaagcacataccacggaaag	2100
Dp	2041	AGTATTTCCAGAGAGACTTGTATCAATAGCGCTGAGAGGTTTAGAGCAKATCMACGMAAAGC	2100
Oy	2101	gaaatctgaaagaagaaaa tggaaagaa cctccaagaagagcccccaagaagtttcatatgctct	2160
Dp	2101	GAATGGAAGAAAGAAATATGAAAGACCTCAAGGAMAGGCCCAAGAACTTTATTAATATGCTTT	2160
Oy	2161	ctgtactaatgtgggtgtagagaagaatccaagaatttttgaagaataatag taagaacaaagatatg	2220
Dp	2161	CTGTACTAGTGGGCTAGAGAAATTCAGATTTTNTTNGAAATATATGTAAAGCAAGATATG	2220
Oy	2221	gacatgtctggcagatctgaagaagaccggagactgatlgtacatgactcttaataataatagtct	2280
Dp	2221	GACATGTGTGGCAGTATGTGAAGAGACCGGACTGATTTTGACATGATCTTTAAATATATAGCTTT	2280
Oy	2281	tgaanaataatagaatccag tgtttataactcttctaataataaataagaatattataaactt	2340
Dp	2281	TGAAATAATNAGATTCACAGCTGTTTATATCTTTCTTTAATAAAAATAGAAATATTTAACCTT	2340
Oy	2341	aaaaaaaaaaaaaaaaaaaaaa 2365	
Dp	2341	AAAAAAAAAAAAAAAAAAAAAAAA 2365	
RESULT	4		
LOCUS	AX086506	2200 bp DNA	
DEFINITION	Sequence 458 from Patent W00112659.	linear	PAT 09-MAR-2001
ACCESSION	AX086506		
VERSION	AX086506.1		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2200)		
	Wiemann, S.		

TITLE Human dna sequences
JOURNAL Patent: WO 0112659-A 458 22-FEB-2001;
German Genome Project (DE)
FEATURES
Source Location/Qualifiers
1. 2200
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 756 a 377 c 523 g 544 t
ORIGIN

Query Match 92.5%; Score 2188.4; DB 6; Length 2200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

176 caacgaacgtcgagcgcgcctctcttggaacatgctccacacgcgaggaagctcccaag 235
1 CACGACGCTCGAGCGGCCCTTCTTGGACATGCTCCACACGAGAGCTCCCAAG 60
236 cctctacgtgggtcgttgcgtacgtcggaagctcgacagtgctccgagcgacagagaa 295
61 CCTCTACGTGGCTCGTTGCTAGTCGCGCAAGCTCGACAGTGTCGCCAGCGCCAGAGGA 120
296 ggcgcgagcgaggaagttgaaatcgaaacaggtccttgaggaataagtgctcgcaagtgctc 355
121 GCGCCGCGGAGAGATTGATCAACAGGCTCTGAGGGAATAGTGTGGCAGAGAGTGCTC 180
356 gctggagaggaacccctcagcccccgagagcgctgagcgcgtggtcagagaaatcgccgc 415
181 GCTGAGAGGACCTCTAGGCCCCCGGAGGCGCTGGCCGCTGGTCAAGAGAACTGCCGC 240
416 tctgttctgcttgaagagcaccttctgttcgcggtgaatcggctggtggtggttaaaaa 475
241 TGTGTTTTGCTTGAAGACCACTTGTGTGGCGGCTAATCGGTCTGTGTGGCTCAAAAA 300
476 taaagaatatacaaaagttacaacaacaccacaatccaataatacaagaacaacagaat 535
301 TAAAGAAATATACAAAGTACACAAACACCAATCCAAATATATCAAGAACAAACAGAAAT 360
536 catlagtcaaaaattcttgcgacgaagcaatcgacaacgaagaagaagaagtgatagaca 595
361 CATTAGTCAAAATTTTGGCAGCAAGGCAATGCAGAAAGCAAAAGCAATGATAGCA 420
596 attctgttaaaaagctcagaagaataatacaatcagaatgcggaattgatactgatacc 655
421 ATTTTGTAAAAAGCTTAGAAGAAATTTCAATTCAGATCGGAATTCGATCGATTC 480
656 aacctctctgtgaaaaaga tggaaagcacaga taaacaatgtctgttcgagagaatcgagcat 715
481 AACCTCTCTGTGGAAAAAGATGGAACACAGATACAGATGTTGTTGCAGAGATCGGCAT 540
716 tgaatagatctggatcaaatctagagaagaagttgaaa tggcaaaaaaacaagtgagcag 775
541 TGATAGATTGGGATCAAAATTAGAGGAAAGCTTTGAATTCGCAAAAAAACAAGTGGCAG 600
776 attacaccaaatctaaagaanaactctlaaagaagctccatgcccacaagctgcacatgca 835
601 ATTATCACCAATTTAAGAAAACTTTTATTAAGAGTCTACCTGCCACAAAGTGCATGTA 660
836 aagtagaagacagatagctggaggaagaanaatttaatacaagtgagatgactgaag 895
661 AAGTAGAAGCAGATAGTTGAGAGAAAGAAATTTTAATATATAGCTGGATGACTTGAAG 720
896 atgggaggaagaagacatccatcccaatctctacgtgcacattgataagccttcaagt 955
721 ATGGGAGAAACAGACTATCCCAATCTTACCTGCACATTTGATGACGCTTTCAAAATG 780
956 atccctgaagttatlgaaaaaatctaaagaagcaggttttcaaaagccaaacactatcagt 1015
781 ATCTGAGCTTATGAAAAACATTAATAAAGGACGTTTCAAAAAGCAACACTATATCGT 840
1016 caaagacatggcccatctgtgtgcagaagaatagatctctatagagtgagcccaagctgaa 1075

841 CACAGCAGTGGCCCATTTGTTGCAAGCAATAGACTTTATAGAGTAGCCCAAGCTGGAA 900
1076 caggaaagaacatctgtctatttaattgcctggatcttaacatcgttccctcaaccagcc 1135
901 CAGAAAGCAATTTGTTATTAAATGCTGATTTATTCATCTGGTCTTCAACCCAGCC 960
1136 ttaaaggtcaaaaggaatagaccggca ttttaatttaactcccatctgggaattggcac 1195
961 TTAAGGTCAAAAGAAATAGACCGCATGTGATTTCACTCCACCTCGGAATTTACAC 1020
1196 ttcaagtagaaggaagatctgtcacaatatcatataaaagggctcagagtgctgtglat 1255
1021 TTCAAGTAGAAGAAATGTTGCAAAATTCATATTAAGGCTTCGAGTGTGTGTAT 1080
1256 atgtgtgtgaaatagatagaataaataaagaagcttaaaaaagtgtagatacataa 1315
1081 ATGGTGTGAATAGATAGATGAACAAATAGAAAGCTTAAAAAGCTGTATATATCATTA 1140
1316 ttgcaactcccgaaagatctgaatctgcaaaatgagtaactctgcaatctgaagata 1375
1141 TTGCAACTCCCGAAGATTGAATGATCTGCAANTGAGTAACCTTGCAATCTGAAGATA 1200
1376 taacctactgtttttagaagaagcagaagaatgttggacatgggattggaacccaga 1435
1201 TAACTACTTGTGTTTATGATGAACAGACAAAGATGTTGACATGGGATTTGAACCCAGA 1260
1436 taaatgaagaattctgttagatgtgcgccagatagcagaacgtttagaacgtgtacat 1495
1261 TAAATGAAGATTTTGTATAGATGTGCCCAATAGGACAGACAGTTATGACAGTGTACAT 1320
1496 ggcctcatcagttatcagcctctgcacaacatctatttgaagaacaaatgactgtctatg 1555
1321 GGCCTATTCAGTTATGATGCTCTGCACAACTTATTTGAAGAACCAATGATTTGTATAG 1380
1556 ttgtacatctgagatcagttctgttgaagtcagtcgaagaataataatctgatacacaag 1615
1381 TTGTGATTTGATGATAGTATGCTGTGATGATGACAGCAAAATATATTTGAACACCG 1440
1616 aggaagaagaatgaggtacatctgaacatcttctcaagaatcttcccaagaagaag 1675
1441 AGCAAGAGAAATGGAGTACATGCAAACTTTCTACAGAGTATGTCATCCACAGCAAAAG 1500
1676 tcaatgtctcgttctcgaagaagctgttgcgataacttaacaaatgacctaactg 1735
1501 TCATTTGCTTCTGTTCTGCAAAAGCTGTGGCAATCACTTCAAGTACCTTAATCTTG 1560
1736 gaaatatcagtaagtcctctgcattgaggaatagaagaagaagaagatcggagaagaagcat 1795
1561 GAAATATATCAGTATAGATCTCTGCATGAGATAGAAACAGAGATCGGAGAAAGCAT 1620
1796 tagagaactttaaaacaggaagaatgagaatlaactaattgcaactgatactagcctcagag 1855
1621 TAGAGAACTTTAAAAAGGCAAGGTGAGAAATCTAATTTGCAATGATCTAGCTTAGAG 1680
1856 gacttgatgtccatgaagtttaacatgctataaatttgaacttccacggaaatatgaag 1915
1681 GACTTGATGTCCATGACGTTTACATGCTATTAATTTGACTTTCACGGAAATTTGAAG 1740
1916 aatacgtacacagaaatagggcgacaggaagaagcaggaagactgtgtctccatataca 1975
1741 AATAGGTACACGAATTTGGGCGCACGGGAAGACAGGAGGACATGCTTTCATTACAA 1800
1976 ctttgactagaagaatgatactggagaggtgtgctctgaaatgataataatctcgaagaagca 2035
1801 CTTTGACTTAGAAATGATTTGAGGGTGGCTGTGAATTTATTAATTAATCTGAAAGACAA 1860
2036 atcagaagatctccaggaagactgttatacgaatgcgaagaagttggaagcacatcaacga 2095
1861 ATCAGAGTATTCAGAGAGACTTGTATCAATGCTGAGAGGTTTGAGAGCACAACAGGA 1920
2096 aaagggaaaatggaaaagaaaatggaaaagcctcaaagaagcccaagaagttcatlaac 2155
1921 AAAAGGAAATGGAAGAAATGGAAGACCTCAAGAGAAAGCCCAAGAACTTTCATTAAT 1980

Oy	2156	gtccctctacaaagtcgggagctaaagaattttttaaaataagtaagaacgaag	2215
Db	1981	gtcttctgactagctggcgatgagcaatttcacaaattttttgaaattatagtagacacgaag	2040
Oy	2216	tattgacacgcttgccagctaaagaacccgacgacttctactgactcttaaaataa	2275
Db	2041	tattgacacgtgtggcagatgtaagaacccgacgacgactgatttcacgattcgtttaaattata	2100
Oy	2276	gtgcttgaataataagaatccagtgctttatcattctctttaaataaataagaagtta	2335
Db	2101	gtgtttgaaaattatagatccagctggtttatcttctttatataaaattgaaacgatttta	2160
Oy	2336	aacttaaaaaaaataaaaaaaataaaaaa	2365
Db	2161	aacttcaaaaaaaataaaaaaaataaaaaa	2190

RESULT	5	
NCBI	NSMB01719	
ACCDS		
DEFINITION	2200 bp. mRNA, linear	PR1 10-MAR-2004
	HSM801719	
	Homo sapiens mRNA, cDNA DKFZp434H2114 (from clone DKFZp434H2114);	
	complete cds.	
ACCESSION	AL136751	
VERSION	AL136751.1	GI:12053018
KEYWORDS		

ORGANISM

REFERENCE AUTHORS

TITLE Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs
JOURNAL Genome Res. 11 (3), 422-435 (2001)

JOURNAL Submitted (18-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY

COMMENT Clone from S. Wleeman, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wleeman@dkfz-heidelberg.de.

FEATURES
SOURCE

```

gene
    CDS
        /organism="homo sapiens"
        /db_xref="caion:9606"
        /clone="DKFZp434H2114"
        /tissue_type="testis"
        /clone_1ib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
        /dev_stage="adult"
        33..1979
        /gene="DKFZp434H2114"
        33..1979
        /gene="DKFZp434H2114"
        /note="strong similarity to RNA helicases"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="CAB66685.1"
        /db_xref="GI:12053019"

```

polyA_signal
polyA_site
BASP COUNT 756 a 377 c 523 g 544 t
Origin

Query Match		92.5%	Score 2188.4	DB 9	Length 2200
Best Local Similarity		100.0%	Pred. No. 0		
Matches 2189		Conservative	0	Mismatches	1
				Indels	0
				Gaps	0
QY	176	caacgaacg	lcggaacg	cgccctctctgtgaaacatcgtccacacgaagagctcccaag	235
Db	1	CAACGACGTCGAGCGCCCTCTTGTGGAAACAATGTCCACACGAGACCTCCAAAG			60
QY	236	cctctacg	tgggtcgtgtgtctgaatctgggaacgtctgcaatgtctccgaacgccaagaagga		295
Db	61	CCTCTACGTGGGTGTTGCTGCTAGTCGGCCAAAGCTTGACAGTGTCCGACGGCAAGAGGA			120
QY	296	gacccgagcgaagagttgnaatcgaacaag	lccctggagatagatgtgtcggcagagtggtc		355
Db	121	GGCCCGGCGAGAGATTGAATTCGAACAGGTCTCTGAGGATATATGTGTGGCAGAGGTGTGC			180
QY	356	gcttggaagaggaacctctgaagcccccggagagcgccgtgacgcctgtgacaaggaatctgcgc			415
Db	181	GCTGAGAGGCGACCTCTAGGCCCCCGAGAGCCGTGGCCGTGTGTGTCACAGGAATGTCCGC			240
QY	416	tgtgttctgtcttgaagagccacttgtcttgccgagtgaaatcgtgtcgtgtgtgtgtcaaaaa			475
Db	241	TGTGTTTTGCTTTGAGAGCGCACTTGTGTGGCGGGTATTCGGTGTGTGTGTGTCAAAA			300
QY	476	taaagaatatacaaaagtaacaaacacccaatactccaataataataagaacaaacgaat			535
Db	301	TAAAGATATACAAAGTACACAAACACCAATTCGAATATACAAAGAACACCAAGAT			360
QY	536	catgaatcaaaaatttttgcgcgaaggaacatgcaacgaagaaacaaagcaagtgaatga			595
Db	361	CATTAGTCAAAATTTTGGCACCAAGCAATGCAAAGCAAAACAAACAGTGTATGACA			420
QY	596	attctgtttaaanaagctagaagaanaattcaattcaagaatgggaattgtatctgtcatcc			655
Db	421	ATTTTGTTTAAAAACCTTGAGAGAAATTTCAATTTCAGATTCGGAAATTCATCTCATATCC			480
QY	656	aaaccttcgtctgnaaagatgnaagacacaga	laacaatgtgtgtgcagagatacggccat		715
Db	481	AACCTTCTGTTGGAAAGATGTGAGACACAGATTAACATGTTGTTCAGGAATCGGCCAT			540
QY	716	tgaatgatctggatccaatctagaagaggaag	cttgaatgtggcaaaaacaaagtgggcag		775
Db	541	TGATTGATTGGGATTCAAATTAGAGAGAAAGTTTGAATGTGCAAAAACCAAGTGGGAC			600
QY	776	attaccacccaattaaagaanaacttlaaagat	lccactgtccacaagtgcacgtgtcaa		835
Db	601	ATTATACCCACCAATTAAAGAAAACCTTTTAAAGAGTCCACTGTCCACAAAGTCCATGTCA			660
QY	836	aagtagaagcagatgaatgtgaggaagaagaanaattlaaataaagctggagataacttgaag			895
Db	661	AAGTGAAGACGATAGTGTGAGAGAAAGAAATTTTAAATATACGTGGGATACCTTGAAGG			720
QY	896	atggggagaaacgaactatcccaat	lctactctgacatttgatgaagccttcaatglt		955
Db	721	ATGGGGAAGAAACGACTATCCCAATCTCTACTCTCAATTTGATGATACCCCTTTCATGTT			780

Qy 956 atccgaggtatctgaaacacataaaagcaggtcttcaaaagccaacccattcaat 1015
Db 781 ATCCTGAGTTATGGAATAACATTAAAAAGCGAGTTTCCAAAAGCCACCTATTCAGT 840
Qy 1016 caacagcatgcccattgctgctcaaggaatagatcctatagagtagaccgaactgaa 1075
Db 841 CACAGCATGGCCCATGTGTGTGCAAGGAATAGATCTTATAGAGTAGGCCCAAGCGAA 900
Qy 1076 caggaagaacattgctatattatccctggagttctatcatctgctcccaagcc 1135
Db 901 CAGGAAGACATGTGTTATTTATGCTGCATGATTTATTCATCTGCTTCAACCCAGCC 960
Qy 1136 tcaaggtcacaaggaatagaccggcaatgcttacttcaactcccaactggaaatagaa 1195
Db 961 TTTAAAGTCAAGGAATAGACCCGGCATGTAGTCTTAACCTCCACCTGGGAATTTGCAAC 1020
Qy 1196 tccaagtagaaggaatgcttgcataatccaataatcaaaaggtccggagtgctgctat 1255
Db 1021 TTCAAGTTAGGAAGGAGATGTTGCAATATTTCATTAAGGGCTTCGAGAGTGTGTGAT 1080
Qy 1256 atggtggtggaatacagagatgaaacaaatagaagagcttaaaaaagtgtagataata 1315
Db 1081 ATGCTGGTGAATAGAGATGAAACAATAGGAAGAGCTTAAAAAGGTGTAGATATCATTA 1140
Qy 1316 tgcgaactcccggaagaatgaaatgactcgaatgaaatgaaatgaaatgaaatgaaat 1375
Db 1141 TTGCAACTCCCGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Qy 1376 taactacttgctttagatgaaagcagacaagaatgctggaatgaaatgaaatgaaat 1435
Db 1201 TAACTACTGCTTTTATGATGAAAGACAGACAGATGTTGCAATGGATTTGAAGCCCA 1260
Qy 1436 taatgaagaattctgtagatgctgcccagatagacagacagatgaaatgaaatgaaat 1495
Db 1261 TAAATGAATTTTGTAGATGTGCGCCAGATAGACAGATGATGATGATGATGATGAT 1320
Qy 1496 ggcctcattcagctcagctcagcaactcttcttgaagaagcaactgaaatgaaatgaaat 1555
Db 1321 GGCCTCATTTCAGTTTCATGCGCTCGCAATCTTATTTAAAGAACCAATATTTGTATAG 1380
Qy 1556 ctgtgacattggaatcagctgctgctgctgctgctgctgctgctgctgctgctgctg 1615
Db 1381 TTGGTACATTTGGATCTAGTTGCTGTAAGTTCACTCAACCAAAATATTTGTAAACACCG 1440
Qy 1616 aggaagagaatcagagtcacacatgcaactcttctcaagagatgcttcaacagagaaga 1675
Db 1441 AGGAAGAGAATGAGTCACATGCAAACTTTCTACAGAGATATCTCAGACAGCAAG 1500
Qy 1676 tcaatgctcctgctcctcgaaagcctgctgctgctgctgctgctgctgctgctgctg 1735
Db 1501 TCATTGCTTGTGTTTCGAAAAGCTGTGGCGATCACTTATCAAGTGACCTTAACTTGG 1560
Qy 1736 gaaatatacagtagagctcctcagatgagatagagaacagagagatcgggaagaagcct 1795
Db 1561 GAAATATATCAGTAGAGTCTCTCATGAGATAGAGAACAGAGATCGGGAGAAAGCT 1620
Qy 1796 tagagaactttaaaacagcgaatgagaatacctaattgcaactgactagcctcagaag 1855
Db 1621 TAGAGAACTTTAAAAAGCGCAAGTAGATCTTAATGTCAACTGATCTTAGCCTTAGAG 1680
Qy 1856 gactgagtcacagacgtctacacatgcttcaatattggaattccacaggaatattgag 1915
Db 1681 GACTTGATGTCATGAGAGTACACATGCTTAATTTTGACTTCCACGGAATATTGAAAG 1740
Qy 1916 aatacgtacacagaaatagggcgcaaggagaagcagaggaagactggttccatcaaa 1975
Db 1741 AATACGTACACCGAATAGGGCGACGGGAAGACGAGGAGACTGGTGTTCATTAAACA 1800
Qy 1976 ctctgactagaatgactgagaggtgctcctcgaattgatatatactcgaagaagcaa 2035
Db 1801 CTTGACTAGAAATGATTTGAGAGGTTGCCTGTGAATTGATTATTTCTGGAAGAGCA 1860

Qy 2036 atcagaatctccagaagagctgtgatatcaatagctcgtgagaggtttgagcacaataagga 2095
Db 1861 ATCAGAGTATTCAGAGAGGCTTGTATCAATAGCTGAGAGCTTGGAGCACAATCAACGGA 1920
Qy 2096 aaaggaaatcgaagaaataatcgaagagaccctcaaggaagggcccaagaagttcattaat 2155
Db 1921 AAAGGAATATGGAAGAAATAATGAAAGACCTCAAGGAAGGCCCAAGAAAGTTTCATTAA 1980
Qy 2156 gctctcgtactagtgaggtagaagatccaagatttttagaataatagtagaagcaag 2215
Db 1981 GTCTTCTGTACTGTGCTGGGTAGACAAATTCAGATTTTATGAAATATGTAAGACAGAG 2040
Qy 2216 taatgacatgttgacagatagaagaccggaactgattgactgattccttaaaataa 2275
Db 2041 TATTGACATGTTGCGAGATGAAAGACCGAGCTGATTTGACTGATTTTAAATTAATA 2100
Qy 2276 gtgttgaaataatagaatccagctttttaaactcttcttaataaaataagaatattta 2335
Db 2101 GTGTTTAAATATATACATCCAGTGTTTTATACTTTCTTTATTAATAATAGAACTATT 2160
Qy 2336 aacttaaaaaaaaaaaaaaaaaa 2365
Db 2161 AACTTGAAAAAATAAAAAAAAAAAAAA 2190
RESULT 6
AC010282
LOCUS AC010282 191132 bp DNA linear PRI 01-SEP-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-353119, complete sequence.
AC010282
AC010282.5 GI:9958012
VERSION
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1. (bases 1 to 191132)
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3. (bases 1 to 191132)
REFERENCE
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Sep 1, 2000 this sequence version replaced gi:9256180.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Piprap Quality >=40 99.2% of Sequence;
Estimated Total Number of Errors is 1.5.
STS Content:
SHGC-100210 G55156
WT-3394 G02870
WT-11812 G24004
SHGC-105345 G58760.
FEATURES
source location/Qualifiers
1. 191132
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-553119"
BASE COUNT 61612 a 37225 c 35694 g 56601 t
ORIGIN
Query Match 39.6%; Score 937.4; DB 9; Length 191132;

Best Local Similarity 73.2%; Pred. No. 2,1e-190;
Matches 1443; Conservative 0; Mismatches 356; Indels 172; Gaps 12;

```
QY 211 tccaccacgaggaagctcccaagcctcagctgagctgctgctgctgaggaactcg 270
    || || || || || || || || || || || || || || || || || || || || ||
Db 82092 TCTAAACAAAAGTAAAGACCAATGCCATTACTGGATCTGCTATCTGGCAACTCA 82151

QY 271 acagtcgccgagcgccagagagagcgcgagagagagagagagagagagagagag 330
    || || || || || || || || || || || || || || || || || || || || ||
Db 82152 ACAGTATCTCGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 82211

QY 331 ggaataagctgctcagagagagagagagagagagagagagagagagagagagag 390
    || || || || || || || || || || || || || || || || || || || || ||
Db 82212 GGATATAGGGGTGGCAGAAAGTGGGAGACCAAGACCCCTCTTA -AACCCGAGAGCTGTG 82270

QY 391 gccgcctgctcagagagagagagagagagagagagagagagagagagagagag 450
    || || || || || || || || || || || || || || || || || || || || ||
Db 82271 GCCGCTACTCTGGAGCCAGCTCTCTTCGATTTGAAGAAAGAGAGAGAGAGAGAG 82330

QY 451 gtaacgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 510
    || || || || || || || || || || || || || || || || || || || || ||
Db 82331 GTGATCACTCAAGGCTGGCTCAAAAATTAAGATTAAGCAACCAACCAACCAACCAATTC 82390

QY 511 caaat-aatcacagaacaacacagagatcagagagagagagagagagagagagag 569
    || || || || || || || || || || || || || || || || || || || || ||
Db 82391 CAGATTAAGAAAAGAGACATCTGGAATCATTAAGTCAACTTTTGGCAGCAAGCGATGCA 82450

QY 570 aacgaagcaaaagcagtgatagagagagagagagagagagagagagagagagag 629
    || || || || || || || || || || || || || || || || || || || || ||
Db 82451 AACAAAGCAAAAGTACTATATAGATTAATGCTTTTAA-----AAAACAAATTAATTCATTC 82507

QY 630 agaatgcggaatctgacagcagcagcagcagcagcagcagcagcagcagcag 689
    || || || || || || || || || || || || || || || || || || || || ||
Db 82508 AGAATACAGAAATGAT----- 82523

QY 690 caatgtctgacagagatcgccatctgagagagagagagagagagagagagag 749
    || || || || || || || || || || || || || || || || || || || || ||
Db 82524 ----- 82523

QY 750 gaaatgagcaaaaacaagctggcgagatccaccacaaatcagaagaactctataaga 809
    || || || || || || || || || || || || || || || || || || || || ||
Db 82524 -----TTAAAMACCAATTAAGAAAACCTTTTAAACA 82556

QY 810 gccacatgccacaagctgcacatgcacaaagtagaagcagatagctggagagagagag 869
    || || || || || || || || || || || || || || || || || || || || ||
Db 82557 GTCAA---CAACAGATTCAATATCCCAAGTCAGACAGACAGACAGAGAGAGCAATTT 82613

QY 870 taataaagctggagagagagagagagagagagagagagagagagagagagagag 929
    || || || || || || || || || || || || || || || || || || || || ||
Db 82614 TAATATATATGATGATAGCTGAATAATGTAAGAAAGCTCTATCCCAATCTATGTTG 82673

QY 930 caacattgatacgccttcacatctcagagagagagagagagagagagagagagagag 989
    || || || || || || || || || || || || || || || || || || || || ||
Db 82674 CACATTTGACGATGCTTTCAATGTTGTTCTCAATGTAATGGAATAATTTAAAGAGAC 82733

QY 990 ttcttaaaagccaaacacactatcagtcacagagagagagagagagagagagagag 1049
    || || || || || || || || || || || || || || || || || || || || ||
Db 82734 TTTTCAAAAACCAACACTATATGCTGACAGGACATGSCCAATCTCTGCAAGAGAAAATA 82793

QY 1050 tcttaagagtagcccaagacagagagagagagagagagagagagagagagagag 1109
    || || || || || || || || || || || || || || || || || || || || ||
Db 82794 TCTTACAGGGGTACCCCAAACTGCTACAGAGAAACATTATGCTATTTAATGCTGATA 82853

QY 1110 tatcatctgctccttaaccagccttaagagagagagagagagagagagagagagagag 1169
    || || || || || || || || || || || || || || || || || || || || ||
Db 82854 TATCATCTGGACTTCCACACACA-----CAGAGCAAAAAGAGATAGACCC-GCATGTTTAT 82908

QY 1170 tctaacctccacctcggaatctgacatcagtagaagagagagagagagagagagag 1229
    || || || || || || || || || || || || || || || || || || || || ||
Db 82909 CCTTATGCCCACTCAGAGAGCTAATCTTTAAGTGAACCTGAGTGAGTCA----- 82958

QY 1230 taagagctcagagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1289
    || || || || || || || || || || || || || || || || || || || || ||
```

```
Db 82959 -ATAAAGTTTAGACATTTGTATACGGTGTGTGAGATGAGACAAACAAATAGAGA 83017

QY 1290 gcttaaaaagagctgatatatatatctgaacatcccgagagagagagagagagagagag 1349
    || || || || || || || || || || || || || || || || || || || || ||
Db 83018 CTTTAGAAGAGAGGTGTGATATATATATGGAACCTCCAAAAGACTGAGATCTGCAAT 83077

QY 1350 gagtaacttcgttaactctgaagaatataaccatctgctcttctgaagagagagagag 1409
    || || || || || || || || || || || || || || || || || || || || ||
Db 83078 GAATTAATTTGAGCTGAGAGAGTAACTACTTGTGTTTATGATTAAGAGACAAATAT 83137

QY 1410 gttagacatgagatggaagagagagagagagagagagagagagagagagagagag 1469
    || || || || || || || || || || || || || || || || || || || || ||
Db 83138 GCTGAGATGGGGGTGTAACCCAAATATATAAATTTTATGATGTGCGCCGAGACAG 83197

QY 1470 gcaagacagttaagcagagctctacatggcctcactcagctcagctcagcagacatct 1529
    || || || || || || || || || || || || || || || || || || || || ||
Db 83198 GCAGAAATATATATGACAG-GCAATCTGGCCATATTCAGTTCATTCATCATCTTA 83256

QY 1530 tttagaagacaaatgagctgctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1589
    || || || || || || || || || || || || || || || || || || || || ||
Db 83257 TTTGAAGAGACCCATGCTTGTCTATGCTGTGATCTGATCTGATCTGTTGCCATTAACTTAA 83316

QY 1590 gaagcaaatataatctgtaacccacgagagagagagagagagagagagagagagagag 1649
    || || || || || || || || || || || || || || || || || || || || ||
Db 83317 TGAAGCAAAATATATATGATCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 83376

QY 1650 acagagatgctacacagagagagagagagagagagagagagagagagagagagagag 1709
    || || || || || || || || || || || || || || || || || || || || ||
Db 83377 AGATCTTATTTCCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 83436

QY 1710 tcaacttaacagtgacctaactctggaatataatcagtagagagctctcagagagagag 1769
    || || || || || || || || || || || || || || || || || || || || ||
Db 83437 TCATTTATCAAGTGAACCTATCTTGAAGAAATATCATGATGATCTCTCAGTGAATAC 83496

QY 1770 agaacaagagagagagagagagagagagagagagagagagagagagagagagagagag 1825
    || || || || || || || || || || || || || || || || || || || || ||
Db 83497 AGAACACAGAGATTTGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 83554

QY 1826 tactaactgcaactgatacagctctcagagagagagagagagagagagagagagagag 1885
    || || || || || || || || || || || || || || || || || || || || ||
Db 83555 TATTATATGCTACTGACATATACATCTGAGCTCTGATGATGATGATGATGATGATGAT 83614

QY 1886 ataacttgcactcctcagagagagagagagagagagagagagagagagagagagagag 1945
    || || || || || || || || || || || || || || || || || || || || ||
Db 83615 ATATTTACATTTTCCATGGAACACTGAAAATATGTAACCTGATATAGGCTGACAGGAA 83674

QY 1946 gagcagagagagagagagagagagagagagagagagagagagagagagagagagag 2005
    || || || || || || || || || || || || || || || || || || || || ||
Db 83675 GAGCTGGAAGAGACTGAGAGGCCAATTAACCTGATGATGATGATGATGATGATGATGAT 83734

QY 2006 ctgaattgataatctcagagagagagagagagagagagagagagagagagagagagag 2065
    || || || || || || || || || || || || || || || || || || || || ||
Db 83735 CCAATATGATGATATTTTGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 83794

QY 2066 tgcctgagagagagagagagagagagagagagagagagagagagagagagagagagag 2125
    || || || || || || || || || || || || || || || || || || || || ||
Db 83795 TGCTTAAGAGACAGAGGAGCACTCAAC-----AATCTAAAAGCAATGGAAGAGAC 83844

QY 2126 ctcaagagagagagagagagagagagagagagagagagagagagagagagagagagag 2176
    || || || || || || || || || || || || || || || || || || || || ||
Db 83845 TTCAAGAGAGCCCAAGAGAGTTTATTAAGATCTTTTATTAAGAAAGTGTTA 83895
```

RESULT 7
AC108109/c 170569 bp DNA linear HTG 25-JAN-2002
LOCUS
DEFINITION Homo sapiens chromosome 5 clone Rpl1-263N24, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
ACCESSION AC108109
VERSION AC108109.1 GI:1836995
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.

REFERENCE 1 (bases 1 to 170569)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170569)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 507693
Center clone name: RPCT-11_263N24

Summary Statistics
Consensus quality: 169426 bases at least Q40
Consensus quality: 169997 bases at least Q30
Consensus quality: 170132 bases at least Q20
Estimated insert size: 187000; agarose-fp estimation
Estimated insert size: 170469; sum-of-ctrls estimation
Quality coverage: 9.04 in Q20 bases; agarose-fp estimation
Quality coverage: 9.92 in Q20 bases; sum-of-ctrls estimation.
NOTE: This is a "working draft" sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1053: contig of 1053 bp in length
* 1054 1153: gap of unknown length
* 1154 170569: contig of 169416 bp in length.
Location/Qualifiers
1. 170569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-263N24"
/clone_11b="RPCT human BAC library 11"

FEATURES

source

BASE COUNT 49143 a 32449 c 33233 g 55643 t 101 others
ORIGIN

Query Match

Best Local Similarity 39.5% Score 934.2; DB 2; Length 170569;
Beat Local Similarity 73.1% Pred. No. 9,9e-190;

Matches 1441; Conservative 0; Mismatches 358; Indels 172; Gaps 12;

QY 211 tccaccaacgagagagctcccaaggcctcagctgagctgctgctgctgagcgaagctcg 270

Db 163746 TCTAACCAAAAGTAGGAGCCCAATGCTATCTGCGTCAATCTGCGCAAGCTCA 163687

QY 271 acaagtgccagagccagagagagagccgagagagagctgagctgagcagagctgag 330

Db 163666 ACAGATCTCTGAGCACCAGAGAGAGACTGCGACAGAGAACTGAATCAAGAGCTTGTA 163627

QY 331 ggataatgctcgagagagctgctgctgagagagacctctagcccccggagagcgag 390

Db 163626 GGATATATGGGGTGGCAAGAGTGGAGACCGAAGAGACCCTCTA-AACCCCGAGCGTGTG 163568

QY 391 gccgctgctcagagaaactgcgcgtctgcttcttctgagagagcacttctgagcg 450

Db 163567 GCCCGTACTACACGAGGACGACTGCTCTGTTTCGATTGAAGAACACTTAACTTGGTGTG 163508

QY 451 gtaatcgctcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 510

Db 163507 GTGATCAGTCACGCTGGTGGTCAAAATAAGATAGCCAAAGACAAACAACATCAAAATC 163448

QY 511 caaat-aatcaagaagaacaccagatcattagtcacaaattttgagcgaaggcaatgca 569

Db 163447 CAGATTAAGAAAGAGACATCTGGAAATCATTAGCAACATTTTGGCAGCAAGCGATGCA 163388

QY 570 aacgaagaagaagaagctgtagacaatttctgtaaaagcttagaagaagaataatcattc 629

Db 163387 AACAAAGCAAAAGTAGTAATAGATAAAGTGCTTTTAA-----AAAACAAATTAACAATTG 163331

QY 630 agaatgcggaattgtagtactgcatccaaactctctgtggaagaagatgaagcagata 689

Db 163330 AGAACACAGAAATTGAT----- 163313

QY 690 caatgtgtcgagagagatcgccatctgtagattggaatcaaatagagaggaagttt 749

Db 163314 ----- 163313

QY 750 gaaatgcaaaaaaagaatggtgagatttaccccaattagaagaaactttatataaga 809

Db 163314 -----TTAAACCAANTGNAACCTTTTCAAAAGA 163282

QY 810 gtccactgccaagaatgcccaltgcaaaagttagaagcagatgttggaggaagaataatt 869

Db 163281 GTCAA---CAACAACTTCAATATCCAAAGTCAAGCAGCAACCGAGAGCAAAATTY 163225

QY 870 taatataagctgagatgagcttgaagagtgagggaagaagcacttcccaatcctacgt 929

Db 163224 TAAATTAATGTGTGATGACGTGAATAATGTGTAAAGAAAGCTCTATCCCATCTGTGTTG 163165

QY 930 cacatttgaatgagcgtcttcaatgattccctgaggtatggaataacataaaaaagcag 989

Db 163164 CACATTTGACGATGCGCTTTCATATGTTCTGTGAAGTATGGAATAATTTAAAGACAG 163105

QY 990 ttccaaagaagaacacttatcagatcagaagagatgagccatttctgtgcaagaataga 1049

Db 163104 TTTTCAAAACCAACACTATTCGCTACACGACATGCGCATTCCTGTGCAAGGAAGA 163045

QY 1050 tcttatagagtagcagcagctggaacagaagaacatttgttacttaatgcttgatt 1109

Db 163044 TCTTACAGGGGTGAGCCCAAACTGCTACAGGAAACAAATTAATGCTATGCTGTGATA 162985

QY 1110 tatctactggtcctcaccagccttaaaagctcaaaagaaatgagccgcagctagct 1169

Db 162984 TATTCAATCTGAGCTTCCAAACGA-----CAGAGAAAAAGAAATGAGACC-GCATGTTAAT 162930

QY 1170 tctactcctcctgagaaatgacactcaagttagaagagaaatgttgcgaataatcata 1229

Db 162929 CCTTATGCCACTCAGAGAGCTAATCTTTAAAGTGAAGCTGAGTGATGCA----- 162878

QY 1230 taaaggtctcgagctgttctgtatatactgtgtgtgtaaatagagataagaataaga 1289

Db 162879 -ATTAAGATTTTGAAGCATTTGTATACGTGTGTGAGATGAGACAAATAATAGAGA 162821

QY 1290 gctttaaagaagcttagatataatctgaactcccggaagatgtgaatgatctgcaaat 1349

Db 162820 CCTTAAAGAAAGGTGTGATATCATATTTGGAACCTCCAAAGACTGATGATCTGGAAT 162761

QY 1350 gaatcaactgctcaacttgaagaatataaactcctgtgtttagatgaagcagaagaat 1409

Db 162760 GAATTAATCTTGTACACCTGAGAAAGTGAACCTACTGTTTGAATGAAGACAGCAATAT 162701

QY 1410 gtctgaacatggaatttgaacccagataatgaagaatttctgttagatgtgtgcccagatg 1469

Db 162700 GCTTGACATGAGGGGTGTGAACCAAGATTAAGAAAGATTTTGTAGATGTGGCCCAACAG 162641

QY 1470 gcaagacgttaagacaaatgctataatgagcccttaactgaattcagctcgcacactta 1529

Db 162640 GCAGATATTTATGACAG-GCAACTTGGCCATATTCAGTCAATGACATACATCAATCTTA 162582

QY 1530 ttgtgaagaaccaatgtctatgttctgtttagatcattgattgttcgtgaagttcagct 1589

Db 162581 TTTGAAGAGCCCATGCTTGTCTATGTGCTGTACACTGCTGTTGCTATTAAGTTTGA 162522

FEATURES

QY	337	agctcgcaaggaagtgctgcgtctggaagagcgaacttaagccccgcgaagccgctgvcgctg	336	/note="assembly_name:Contig37"	53224..56857	misc_feature
DB	41728	AGGGGACGACGAGAGCAGCTGCTGGAAGGGCCCTTCGGCCATACGAGACCAGACACGA	41787	/note="assembly_name:Contig38"	56958..62597	misc_feature
QY	337	ggtcaaggaagacgcgcgtgctgttctgtcttgaagagcgaacttctgvcggtatc	456	/note="assembly_name:Contig39"	62698..67833	misc_feature
DB	41788	GGCTCCCGTGAACCCACCTCTGCTTTAAATTAAGAACATATGTTGGTGTGTCATT	41847	/note="assembly_name:Contig40"	67934..73053	misc_feature
QY	457	ggtcgtgctggtggtcaaaaataagaatatacaagaatcaaaaacgcgaatccaata	516	/note="assembly_name:Contig41"	73154..78894	misc_feature
DB	41848	GCTTACAGTGAGATCAAAAATATTAAGATCTACACATTTGCACAAACCTTAATTCAGATC	41907	/note="assembly_name:Contig42"	78995..85698	misc_feature
QY	517	atcaagaagaacacccaatcatatgaccaaatcttctgcgcgaagccaatgcaacgaaa	576	/note="assembly_name:Contig43"	85799..91004	misc_feature
DB	41908	ATTAACGGGGAATCTGGAAGCAAAGATCAAAATTTTGGCAATAGGGAATGAAAGCAAAG	41967	/note="assembly_name:Contig44"	91105..94855	misc_feature
QY	577	gcaaaagcagatgataagcaaatctgtcttaaaaagcctagaagaataatcacatctcgaatgc	636	/note="assembly_name:Contig45"	94956..101458	misc_feature
DB	41968	GCCAAAGCGGTATGAAACATTATTATGAAMACAAAGAAAGCTCCAACTCGAATTCACAT	42021	/note="assembly_name:Contig46"	101599..108213	misc_feature
QY	637	ggaattgatacgtgcgtcaactcttgtctgaaaagaatggaagcagataacaatgctt	696	/note="assembly_name:Contig47"	108314..114579	misc_feature
DB	42028	GTCGATATATGCTGCATCTCCCAAAACCCCTTATTTGAGAAATTTAGCGAAATGACATTGT	42086	/note="assembly_name:Contig48"	114680..119328	misc_feature
QY	697	gttcgcsagaatcgcgcgaatgatagtatggtgatacaaatitagaagaaagtttgaatg	756			
DB	42087	-ATGGAGAGTGGAGCATTTGTCAAAATTTGGATGGCATTTAGGCGACAGTGTGAGACTGC	42145			
QY	757	caaaaaaacaagttgagcagaattacacccaatitagaanaaacattataagaatccact	816			
DB	42146	GAANAAGAAATATGCGCAGATCTACACCAAGTTAAGAAAACCTTTTACATAGAAATCCAAA	42205			
QY	817	gccacaagtgcacatgataaagttaagaagcgaatagttgagaagaagaataatctaatata	876			
DB	42206	GCAACAAAGTGTGATCTGAAATAGCAGGTGATTAATCTGGAGAAAGAAATTTTCAACATA	42265			
QY	877	acgttgagatgaacttgaagaaatgggagaaagcaactatcccaatcccaacttgaaatctt	936			
DB	42266	ACGTGTGATGACTTGAAAGATGCTGCAAAAGCTCTCATTTCCAAAACCAACTTGCAGATT	42325			
QY	937	gataagcccttcaatgatatctcgaagttatggaanaaattaaaaagcagtgcttccaa	996			
DB	42326	AAAGAGCGCTTTTACCAATACCTCATTTCTCTGAAAGCAATATATAGGATGGGATGTA	42385			

QY 211 tccaccacgagagagcctcccaagccttaagcttggtgctgctgagtcggcgaaagctcg 270
DB 91207 TCTAACCAAAAAGTAGAGCCCAATGCTTATCTGGGTC -ATGCTATCTGCCCAACTCA 91149
QY 271 acagctgcccagagccagagagagcgagcgagaggttgatctgaacaggtctccag 330
DB 91148 ACAGTATCCTGAGCACACGAGAGAGACTGGCAGAGAACTGATCAAGAGGT -GTGAA 91090
QY 331 ggaatagctgagagaggtggtcgctgagagagagcactctagcccccgagagcgctg 390
DB 91089 GGAATATAGGGGTGGCAGAGAGTGGGAGACCCCTCTA -AACCCGGAGGCTGTG 91031
QY 391 gcccgtgctcagagagagctgcgctgcttgctcttgaaagagcactctgttggcgcg 450
DB 91030 GCGGCTAGTCACTGGAGCCACTGCTCTGTGGATGGAAGACGACTAGTTGGTGTG 90971
QY 451 gtaatcgctgctgctgagctcaaaaataaagaatataaagatacaacaacaccacaatc 510
DB 90970 GTGATCAGTCACGGTGGGTCAAAAATTAAGATACCCAAACACACAAACATCAAAATC 90911
QY 511 caaat-aataaagaagaacacacgaatcaatctagtcaaaaatttggcagcaggaatgca 569
DB 90910 CAGATTAAGAAAAGAGACATCTGGAATCATTAAGTCAAACTTTTGGCAGCAAGCGATGCA 90851
QY 570 aacgaagaacaaagcagctgagatacaaatcttgtaaaagctagaagaataatactac 629
DB 90850 AACAAAACCAAACTAGTAATTAATTAATGCTTTTAA---AAAAACAATTAACATTC 90794
QY 630 agaatcgagaaatctgatactgcaatccactctgcttgaaagaatgaaagacagataa 689
DB 90793 AGAACACAGAAATGAT----- 90776
QY 690 caatgtctgctgagagagatcgccatctgatactggagataaattgagagagaagttt 749
DB 90777 ----- 90776
QY 750 gaaatgagcaaaaaaagaatgagagattacaccaaataagaanaactttataaga 809
DB 90777 -----TTAAAMACCAATTCAGAAAACCTTTTACAAAGA 90745
QY 810 gtccactgccaagaatgcaatgcaaaagtagaagataagttggagagaaagaataatt 869
DB 90744 GTCAA---CAACAAGTTCAATATCCCAATGCAACGACAGACACGGAGAGCAAAATTT 90688
QY 870 taatatacgtggagagtaactctgaagaatgggagagaacagactatcccaatcctactg 929
DB 90687 TAATTAATTAATGCTGATGACGTGAATAATGTAAGAACGTCCTATCCCAATCTTAGTGG 90628
QY 930 caaatctgatacgccttcaatcgtatcctgaggtatagaaacaataaagaagcag 989
DB 90627 CACATTTGACGATGCTTTCAATGTTCTTGAGTATTAAGAAATATTAATAAGACACAG 90568
QY 990 ttctcaaaagccaacactatctcaatcagtcagagcagatgcccactgtgttgcaagaata 1049
DB 90567 TTTTCAAAAACCAACACTATATTCGGTACAGCAAGGCAATGCTCTGCAAGGAAAGA 90508
QY 1050 tcttaagagtaagccagagctgaacagagaaagacatgtgtatttaatacctcgagtt 1109
DB 90507 TCTTACAGGGGTAGCCCAAACTGTACAGGAAAGAAACATTAATGCTATTAATGCTCGATA 90448
QY 1110 tattcatctgctctcaacccagccttaagaatgaagaataagaccggacgtttagt 1169
DB 90447 TATTCATCTGCACTTCACAAACA---CAGAGAAAAAAGAAATAGACCC-GCATCTTGTAGT 90393
QY 1170 tctaaactccacctcaggaatctgacacttcaagtagaagagaaatgtctgcaaatatctta 1229
DB 90392 CCTTATGCCCCACTAGAGAGTAACTTTAAGTGAAGCTGACTGATGCA----- 90341
QY 1230 taagagctcctcgagatgttctgtatatacgtgtggaataagataagataagataaga 1289
DB 90342 -ATTAAGTTTAGAACCAATTTGTATACGGTGTGAGATGAGAACAAACAATAGAGA 90284

QY 1290 gcttaaaaaagtgtagatacatataatctgcaactcccggaagaatgtgaatgctgcaaat 1349
DB 90283 CCTTAGGAAGAAGAGAAATATCATATTGGCAACTCCCAAAAGACTGATGATCTCAAAAT 90224
QY 1350 gagaactctgcaactcaatgaagaataataactacttggttgtagaagaagcaagaat 1409
DB 90223 GAATTAATCTTGTCAAGCTGAGAAAGTGAATCTACTTGGTTTATGATTAAGCAGCAATAT 90164
QY 1410 gtggacactggaatttgaacccagataatgaagaatttgtagatgctgcccagatag 1469
DB 90163 GCTGACATGGGTGTGAACCCAAAGATTAATGAATTTGTATANNNNNNNNNNNNNNNN 90104
QY 1470 gcaagcagtatgaccagtgctacatgctcatcactgactgctgcctgcacaatctta 1529
DB 90103 NNN 90044
QY 1530 ttgaaagaaccaatgattgctta 1533
DB 90043 NNNNNNNNNNNNNNNNNNNNNNA 90020

RESULT 10

AC068247

LOCUS

DEFINITION

AC068247

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AC068247 129261 bp DNA linear HTG 09-JUN-2000
Homo sapiens clone RP11-592J5, WORKING DRAFT SEQUENCE, 33 unordered
pieces.
AC068247.2 GI:8389573
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 129261)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-592J5
Unpublished
2 (bases 1 to 129261)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckhalter, B., Brown, A., Burrett, G.,
Campolano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeBartolo, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domingo, M., Doyle, M., Ferrer, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Labrecque, K., Lamazares, R., Landers, T., Lebecky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McKean, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Meneus, L., Milnova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivier, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teste, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, O., Zimmer, A. and Zody, M.
Direct Submission
Submitted (30-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 9, 2000 this sequence version replaced gi:17671314.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/MW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: U7465


```

----- Summary Statistics
Center clone name: 592_J-5
Sequencing vector: M13: M77815: 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 111924 bases at least Q40
Consensus quality: 120625 bases at least Q30
Consensus quality: 124069 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 126061; sum-of-contigs
Quality coverage: 2.7 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1020 1119: gap of 1019 bp in length
1120 2233: contig of 1114 bp in length
2234 2333: gap of 100 bp
2334 3388: contig of 1055 bp in length
3389 3488: gap of 100 bp
3489 4745: contig of 1257 bp in length
4746 4845: gap of 100 bp
4846 7126: contig of 2281 bp in length
7127 7226: gap of 100 bp
7227 9254: contig of 2028 bp in length
9255 9354: gap of 100 bp
9355 10467: contig of 1113 bp in length
10468 10567: gap of 100 bp
10568 12043: contig of 1476 bp in length
12044 12143: gap of 100 bp
12144 13675: contig of 1532 bp in length
13676 13775: gap of 100 bp
13776 15207: contig of 1432 bp in length
15208 15307: gap of 100 bp
15308 17676: contig of 2369 bp in length
17677 17776: gap of 100 bp
17777 20366: contig of 2590 bp in length
20367 20466: gap of 100 bp
20467 23113: contig of 2647 bp in length
23114 23213: gap of 100 bp
23214 25686: contig of 2473 bp in length
25687 25786: gap of 100 bp
25787 28215: contig of 2429 bp in length
28216 28315: gap of 100 bp
28316 30549: contig of 2234 bp in length
30550 30649: gap of 100 bp
30650 33299: contig of 2650 bp in length
33300 33399: gap of 100 bp
33400 36023: contig of 2624 bp in length
36024 36123: gap of 100 bp
36124 37903: contig of 1780 bp in length
37904 38003: gap of 100 bp
38004 41092: contig of 3089 bp in length
41093 41192: gap of 100 bp
41193 45421: contig of 4229 bp in length
45422 45521: gap of 100 bp
45522 48381: contig of 2860 bp in length
48382 48481: gap of 100 bp
48482 54521: contig of 5840 bp in length
54522 54421: gap of 100 bp
54422 59074: contig of 4653 bp in length
59075 59174: gap of 100 bp
59175 62454: contig of 3280 bp in length
62455 62554: gap of 100 bp
62555 66890: contig of 4336 bp in length
66891 66990: gap of 100 bp

```

```

FEATURES
source
*
* 66991 70314: contig of 3324 bp in length
* 70315 70414: gap of 100 bp
* 70415 77680: contig of 7266 bp in length
* 77681 77780: gap of 100 bp
* 77781 82561: contig of 4781 bp in length
* 82562 82661: gap of 100 bp
* 82662 91256: contig of 8595 bp in length
* 91257 91356: gap of 100 bp
* 91357 96466: contig of 8290 bp in length
* 96467 99746: gap of 100 bp
* 99747 115414: contig of 15668 bp in length
* 115415 115514: gap of 100 bp
* 115515 129261: contig of 13747 bp in length.
Location/Qualifiers
1..129261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1bp="RPC1-11 Human Male BAC"
1..1019
/note="assembly_fragment"
1120..2233
/note="assembly_fragment"
2334..3388
/note="assembly_fragment"
3489..4745
/note="assembly_fragment"
4846..7126
/note="assembly_fragment"
7227..9254
/note="assembly_fragment"
9355..10467
/note="assembly_fragment"
10568..12043
/note="assembly_fragment"
12144..13675
/note="assembly_fragment"
13776..15207
/note="assembly_fragment"
15308..17676
/note="assembly_fragment"
17777..20366
/note="assembly_fragment"
20467..23113
/note="assembly_fragment"
23214..25686
/note="assembly_fragment"
25787..28215
/note="assembly_fragment"
28316..30549
/note="assembly_fragment"
30650..33299
/note="assembly_fragment"
33400..36023
/note="assembly_fragment"
clone_end:T7
vector_side:right"
36124..37903
/note="assembly_fragment"
38004..41092
/note="assembly_fragment"
41193..45421
/note="assembly_fragment"
45522..48381
/note="assembly_fragment"
48482..54321
/note="assembly_fragment"
54422..59074
/note="assembly_fragment"
59175..62454
/note="assembly_fragment"
62555..66890
/note="assembly_fragment"

```

```

misc_feature      66991..70314
                  /note="assembly_fragment"
misc_feature      70415..77680
                  /note="assembly_fragment"

Query Match      17.4%; Score 412.6; DB 2; Length 139261;
Best Local Similarity 95.7%; Pred. No. 5.2e-78;
Matches 424; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 26 aacggccgcagtgctgctgaaagtgaggatagaagagctggcgaggagctgag 85
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122477 AATGGCTGCTGAATCACTGGAAGTGGCGGATGACAGCTGGCGGGGGCTAG 122536
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 86 cctcgtagggctccttaagtagcgctgctgagctccctggacagctagctacgac 145
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122537 CCTCGTGGGGCTCCTTAAGTAGCGGCTGGCTTCCCTGGCACAGCTTACGAC 122596
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 146 gtcaagctcagtgctgctgacagctgagcaagacgtcgagcgccctcttgaa 205
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122597 GTCACGGTCAAGTGGTGCAGACCTGGAGGAGACGTCGACGCGCCCTTCTGAA 122656
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 206 caatgtcccaacacaggaagagctcccaagcctctacgtggtctgtagtgcgagaa 265
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122657 CAATGTCCACACAGGAGAGAGCTCCCAAGGCTCTACGTGGGCTGTGCTAGTGGCGAA 122716
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 266 gctcgacagtgctcccgagcgccgaggaagcgcgagagagtgtaactgaacagatc 325
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122717 GCTCGACAGTGTCCGAGGCGGACAGAGAGCGCGGAGAGGATTGAATCGAACAGGTC 122776
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 326 ctgaaggagatagtgtcggcagaagtggtgctgtaggaagacacctctagcccgagag 385
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122777 CTGAGGAGATATAGTGTGGCAGAGGTGGTCCCTGGAAGAGCACCTTAGGCCCCGGAGG 122836
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 386 ccgtagcgctggttcaacaggaagacgtcgtgttctgttgaagagcacattgttg 445
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122837 CCGTGGCCGCTGCTCACAGGAACTGCCGTGTCTTCTTGAAGACCACTTGTGTG 122896
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 446 ggcggtgaatcggtcgtgtgag 468
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122897 GCGCGTAACTGCTGAGAAATGGG 122919
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
AC019205/c 198052 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-398K22 from 6, complete sequence.
DEFINITION AC019205
ACCESSION AC019205.4 GI:12704683
VERSION 1
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 198052)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
2 (bases 1 to 198052)
Nguyen, C., Drone, K., Laplant, Y. and Baum, D.
The sequence of Homo sapiens BAC clone RP11-398K22
Unpublished (2001)
3 (bases 1 to 198052)
Waterston, R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 198052)
Waterston, R.H.
Direct Submission
Submitted (07-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

```

REFERENCE
AUTHORS MO 63108, USA
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 198052)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 7, 2001 this sequence version replaced gi:7630978.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0398K22
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPc1-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catehese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries: Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC068247. Actual start of this
clone is at base position 1 of RP11-398K22; actual end is at base
position 198052 of RP11-398K22.

The sequence from base position 170120 to 170941 is represented by
PCR from clone DNA.
Location/Qualifiers
1..198052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-398K22"
/clone_lib="RPc1-11"
24..325
/rpt_family="Alu"
repeat_region
560..656
/rpt_family="Alu"
repeat_region
665..972
/rpt_family="Alu"
repeat_region
992..1156

```

repeat_region /rpt_family="Alu" 1163. .1388
repeat_region /rpt_family="L2" 1389. .1682
repeat_region /rpt_family="Alu" 1683. .1756
misc_feature /rpt_family="L2" 1991. .2408
EST R08522 (NID:g76045) ye95b11.r1"
repeat_region /note="similar to 2301. .2493
repeat_region /rpt_family="Alu" 2630. .2940
repeat_region /rpt_family="Alu" 3610. .3721
repeat_region /rpt_family="MIR" 4050. .4292
repeat_region /rpt_family="Alu" 4846. .5144
repeat_region /rpt_family="Alu" 5463. .5767
repeat_region /rpt_family="Alu" 5959. .6126
repeat_region /rpt_family="Alu" 6208. .6327
repeat_region /rpt_family="L1" 6381. .6691
repeat_region /rpt_family="Alu" 6752. .7035
repeat_region /rpt_family="Alu" 7216. .7343
repeat_region /rpt_family="L1" 7347. .7428
repeat_region /rpt_family="Alu" 7450. .7527
repeat_region /rpt_family="Alu" 7767. .7902
repeat_region /rpt_family="Alu" 8078. .8284
repeat_region /rpt_family="L1" 8292. .8383
repeat_region /rpt_family="L1" 8409. .8512
repeat_region /rpt_family="Alu" 8515. .8614
repeat_region /rpt_family="L1" 8615. .8804
repeat_region /rpt_family="Alu" 8905. .9203
repeat_region /rpt_family="L1" 9235. .9336
repeat_region /rpt_family="MER2_type" 9333. .9393
repeat_region /rpt_family="MER2_type" 9397. .9661
repeat_region /rpt_family="L1" 9662. .9774
repeat_region /rpt_family="Alu" 9775. .10076
repeat_region /rpt_family="Alu" 10077. .10088
repeat_region /rpt_family="Alu" 10105. .10297
repeat_region /rpt_family="Alu" 10302. .10450
repeat_region /rpt_family="L1" 10458. .10758
repeat_region /rpt_family="Alu" 10494. .10753
misc_feature /note="similar to 10806. .11092
repeat_region /rpt_family="Alu" 11101. .11407
repeat_region /rpt_family="Alu"

EST BE890623 (NID:g10349131),"

repeat_region 11439. .11725 /rpt_family="Alu"
repeat_region 11728. .11916 /rpt_family="L1"
repeat_region 11949. .12059 /rpt_family="L1"
repeat_region 12060. .12371 /rpt_family="L1"
repeat_region 12372. .12406 /rpt_family="Alu"
repeat_region 12431. .12740 /rpt_family="L1"
repeat_region 12745. .12898 /rpt_family="Alu"
repeat_region 12899. .13211 /rpt_family="L1"
repeat_region 13212. .13612 /rpt_family="Alu"
repeat_region 13724. .14014 /rpt_family="L1"
repeat_region 14341. .14655 /rpt_family="Alu"
repeat_region 15327. .15489 /rpt_family="Alu"
repeat_region 15534. .15738 /rpt_family="Alu"

Query Match 17.4% Score 411. DB 9: Length 198052;

Best Local Similarity 95.5% Pred. No. 1.2e-77;

Matches 423; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 26 aacgcccagtcgtgctggaagagtgagcagagtagagacgtgagcggggggtacg 85
DB 48837 AATGCTGCTGCTGAATCATGCGAGTGGCGGATGAGAGCGTGGCGGGGCTAG 48778
QY 86 cctgtgcccgccttaagtagcggcgtgctgcttccctgagcgccttaacgac 145
DB 48777 CCTGTGCGGGCTCTTAAGTAGCGCGCTGCTGCTTCCCTGACAGCTTACGAC 48718
QY 146 gtcaagtcagtgctgacagagctggaagcgaacgactggacgcgcctcttgaa 205
DB 48717 GTCAAGCTCAGTGTGCTGAGACTGAGAGCGCAAGCGACGCTGCGCCCTCTTGAA 48658
QY 206 caatgtcccaacagagagagctcccaagcctctcgtgggtcttgtagtcggaa 265
DB 48657 CAATGTCCACACAGAGAGAGCTCCAAAGCCTCTTACGTGGTCTTGTCTATGCGCA 48598
QY 266 gctgacagtgctccgaagcgcacagagagagccgagcggagagttgaatcgaaacagtc 325
DB 48597 GCTGACAGTGTCCCGAGCGCCAGAGAGAGCGCGCGAGGAGTGAATCAACAGGTC 48538
QY 326 ctgagagataatgtgctgcagaagtgctgtagagagacacctagagcccgagg 385
DB 48537 CTGAGGATATATGTGCGGACAGAGTGTGCTGAGAGGACACTTACGCCCGGAGG 48478
QY 386 ccgtgagcgcgtgctcaagagaactgcgcgtgtgttcttgtaagagcaacttgctg 445
DB 48477 CCGTGGCGCGCTGTCTACAGAGAACTGCCGCTGTGTTGCTTGAAGGCACTTGTG 48418
QY 446 ggcgggaatcgctgctgtgtgg 468
DB 48417 GCAGGATATGCTGTGAATGCG 48395

RESULT 12
AC010091/c 5686 bp DNA linear PRI 30-SEP-2000
LOCUS Homo sapiens BAC clone RP11-295A1 from 7q11.23-q21.1, complete
DEFINITION
ACCESSION AC010091
VERSION AC010091.2 GI:9211377
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 56864)

REFERENCE Sultston, J.E. and Waterston, R.
Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792

REFERENCE 2 (bases 1 to 56864)
Cordum, H., Stoneking, T. and Glaser, E.
The sequence of Homo sapiens BAC clone RP11-295A1
JOURNAL Unpublished
TITLE

REFERENCE 3 (bases 1 to 56864)
Waterston, R.H.
Direct Submission
Submitted (11-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 56864)
Waterston, R.H.
Direct Submission
Submitted (15-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 5 (bases 1 to 56864)
Waterston, R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 15, 2000 this sequence version replaced gi:5870309.
COMMENT ----- Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH025A01

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRF/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACpac.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-5B9, 200 bp overlap; the

Clone sequenced to the right is RP5-1098B1, 200 bp overlap. Actual start of this clone is at base position 91481 of RP11-5B9; actual end is at base position 59117 of RP5-1098B1.
Location/Qualifiers

FEATURES

source

repeat_region	3..631	/rpt_family="MER1_type"	25..38	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"
misc_feature	25..38	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"	25..38	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"
misc_feature	25..38	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"	25..38	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"
repeat_region	652..879	/rpt_family="MER1_type"	652..879	/note="similar to EST AA716347 (MID:g2728621) zh29d04.s1"
repeat_region	883..1241	/rpt_family="MER1_type"	883..1241	/rpt_family="MER1_type"
repeat_region	2702..2958	/rpt_family="L2"	2702..2958	/rpt_family="L2"
repeat_region	2959..3075	/rpt_family="MER2_type"	2959..3075	/rpt_family="MER2_type"
repeat_region	3076..3128	/rpt_family="Alu"	3076..3128	/rpt_family="Alu"
repeat_region	3129..3297	/rpt_family="(CA)n"	3129..3297	/rpt_family="(CA)n"
repeat_region	3298..3365	/rpt_family="Alu"	3298..3365	/rpt_family="Alu"
repeat_region	3511..3808	/rpt_family="MER2_type"	3511..3808	/rpt_family="MER2_type"
repeat_region	4452..4841	/rpt_family="Alu"	4452..4841	/rpt_family="Alu"
repeat_region	5091..5183	/rpt_family="MALR"	5091..5183	/rpt_family="MALR"
repeat_region	5389..5530	/rpt_family="MALR"	5389..5530	/rpt_family="MALR"
repeat_region	5849..6295	/rpt_family="L2"	5849..6295	/rpt_family="L2"
repeat_region	6611..8631	/rpt_family="MALR"	6611..8631	/rpt_family="MALR"
repeat_region	8859..8883	/rpt_family="L2"	8859..8883	/rpt_family="L2"
repeat_region	9379..9668	/rpt_family="AT-rich"	9379..9668	/rpt_family="AT-rich"
repeat_region	10294..10473	/rpt_family="L2"	10294..10473	/rpt_family="L2"
misc_feature	10379..10473	/note="similar to EST AI582192 (MID:g4568089) tq63e11.x1"	10379..10473	/note="similar to EST AI582192 (MID:g4568089) tq63e11.x1"
misc_feature	10518..10896	/note="similar to EST AI871836 (MID:g5545885) vm51h12.x1"	10518..10896	/note="similar to EST AI871836 (MID:g5545885) vm51h12.x1"
misc_feature	10518..10901	/note="similar to EST AI582192 (MID:g4568089) tq63e11.x1"	10518..10901	/note="similar to EST AI582192 (MID:g4568089) tq63e11.x1"
misc_feature	10539..10896	/note="similar to EST AI871836 (MID:g5545885) vm51h12.x1"	10539..10896	/note="similar to EST AI871836 (MID:g5545885) vm51h12.x1"
misc_feature	10644..10891	/note="similar to EST AI189368 (MID:g3740577) qd05g04.x1"	10644..10891	/note="similar to EST AI189368 (MID:g3740577) qd05g04.x1"
misc_feature	10644..10897	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"	10644..10897	/note="similar to EST AI241674 (MID:g3837071) qu70d11.x1"
repeat_region	11201..11565	/note="similar to EST AA716347 (MID:g2728621) zh29d04.s1"	11201..11565	/note="similar to EST AA716347 (MID:g2728621) zh29d04.s1"
repeat_region	12448..12555	/rpt_family="L1"	12448..12555	/rpt_family="L1"
repeat_region	13589..13730	/rpt_family="AT-rich"	13589..13730	/rpt_family="AT-rich"
repeat_region	13762..13805	/rpt_family="MER1_type"	13762..13805	/rpt_family="MER1_type"
repeat_region	13940..13979	/rpt_family="(TTAAT)n"	13940..13979	/rpt_family="(TTAAT)n"
repeat_region	14039..14859	/rpt_family="AT-rich"	14039..14859	/rpt_family="AT-rich"
repeat_region		/rpt_family="MER1_type"		/rpt_family="MER1_type"

```
repeatL_region 15600. .16114 /rpt_family="MER1_type"
repeatL_region 16171. .16501 /rpt_family="MER1_type"
repeatL_region 17358. .17955 /rpt_family="MER2_type"
repeatL_region 18111. .18340 /rpt_family="MER2_type"
repeatL_region 18487. .18515 /rpt_family="AT_rich"
repeatL_region 18628. .18965 /rpt_family="Retroviral"
repeatL_region 19221. .19352 /rpt_family="MIR"
repeatL_region 19815. .19974 /rpt_family="L2"
repeatL_region 20255. .20519 /rpt_family="L2"
repeatL_region 20909. .22180 /rpt_family="Mariner"
repeatL_region 22181. .22200 /rpt_family="TG)n"
repeatL_region 22211. .22395 /rpt_family="Mariner"
repeatL_region 23854. .23931 /rpt_family="AT_rich"
repeatL_region 25556. .25630 /rpt_family="MER94"
repeatL_region 26415. .26455 /rpt_family="(TGAA)n"
repeatL_region 26562. .26843 /rpt_family="Alu"
repeatL_region 26847. .26900 /rpt_family="(TAA)n"
repeatL_region 26985. .27117 /rpt_family="MER1_type"
repeatL_region 27433. .27561 /rpt_family="MIR"
repeatL_region 28264. .28530 /rpt_family="MELR"
repeatL_region 31086. .31174 /rpt_family="AT_rich"
repeatL_region 32846. .33080 /rpt_family="MIR"
repeatL_region 33877. .33948 /rpt_family="MELR"
```

[illegible][illegible]

Db 61 ATTAAAGAAAACTTTTATTAAGAGTCCACTGCCACAGTGCATGTCAAAAGTAGAAGCA 120
Qy 847 gatagtctggagaaagaataatttaataataacgttggaatgaacttgaaggaatgggagaa 906
|||||
Db 121 GATAGTGGAGGAAGAAATTTTAATATACGTGGATGACTTGAAAGATGGGAGAA 180
Qy 907 cgaactatcccccaatcctacccgacatttgatgaagccttcaatgttactcctgaagtt 966
|||||
Db 181 CGACCTATCCC--AATCTACCTGCACATTTGATGACGCCCTTCAATGTATCCTGAAGGTT 238
Qy 967 atggaacacataaagaagcaggcttccaagaagccacactatcagtcacagcagatg 1026
|||||
Db 239 ATGGAACACATTAAAGGCGAGTTTCAAAAGCCACACCTATTCACTCAGTCACAGGATGG 298
Qy 1027 cccattgtgttgcagaaggaatagatc 1051
|||||
Db 299 CCCATTGTGTTCAGAGATATGATC 323

RESULT 15

AR173200

AR173200

323 bp

DNA

linear

PAT 17-DEC-2001

LOCUS

AR173200

Sequence 1 from patent US 6303756.

DEFINITION

AR173200

AR173200.1 GI:17912691

ACCESSION

AR173200.1

GI:17912691

VERSION

AR173200.1

GI:17912691

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unknown.

REFERENCE

1 (bases 1 to 323)

AUTHORS

Martelange V, De Smet C, and Boon-Faljeur T.

TITLE

Tumor associated nucleic acids and uses therefor

JOURNAL

Patent: US 6303756-A 1 16-OCT-2001;

FEATURES

Location/Qualifiers

source

1..323

BASE COUNT

120 a 55 c 72 g 76 t

ORIGIN

/organism="unknown"

Query Match 13.08; Score 307.8; DB 6; Length 323;
Best Local Similarity 98.84; Pred. No. 1.1e-55;
Matches 321; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 727 gatcaaatagagaggaaggttgaatgtgcaaaaaaagaatgggcagattaccacca 786
|||||
Db 1 GATCAAAATTAGAGAGGAAGGTTGGAATGCAAAAAACAAGTGGGACGATTACCA 60
Qy 787 attaagaaaaaactttataaagagtcgaactgcacaagtgccaatgtcaaaatagaagca 846
|||||
Db 61 ATTAAAGAAAACTTTTATTAAGAGTCCACTGCACACAGTGCATGTCAAAAAGTAGAAGCA 120
Qy 847 gatagtctggagaaagaataatttaataataacgttggaatgaacttgaaggaatgggagaa 906
|||||
Db 121 GATAGTGGAGGAAGAAATTTTAATATACGTGGATGACTTGAAAGATGGGAGAA 180
Qy 907 cgaactatcccccaatcctacccgacatttgatgaagccttcaatgttactcctgaagtt 966
|||||
Db 181 CGACCTATCCC--AATCTACCTGCACATTTGATGACGCCCTTCAATGTATCCTGAAGGTT 238
Qy 967 atggaacacataaagaagcaggcttccaagaagccacactatcagtcacagcagatg 1026
|||||
Db 239 ATGGAACACATTAAAGGCGAGTTTCAAAAGCCACACCTATTCACTCAGTCACAGGATGG 298
Qy 1027 cccattgtgttgcagaaggaatagatc 1051
|||||
Db 299 CCCATTGTGTTCAGAGATATGATC 323

